

Developing multivalent nanoparticle vaccines against current and future viruses

*Thesis by*  
*Alexander Cohen*

In Partial Fulfillment of the Requirements for the degree of  
Doctor of Philosophy

Caltech

CALIFORNIA INSTITUTE OF TECHNOLOGY

Pasadena, California

2021

(Defended May 13, 2021)

© 2021

Alexander Cohen

ORCID: 0000-0002-2818-656X

## ACKNOWLEDGEMENTS

(intentionally left blank)

## ABSTRACT

The 1918-1919 flu pandemic resulted in an estimated 50 to 100 million deaths worldwide, making it the deadliest pandemic in modern history. It was caused by a new influenza virus that likely spilled over from birds and reassorted with a human influenza virus. Since the human population was immunologically naïve to this virus, transmission and lethality was much higher than for seasonal influenza outbreaks. Numerous pandemic influenza viruses emerged within the next century, with none causing the same amount of carnage. There is likely to be future influenza pandemics, with wild migratory birds being carriers of a wide swath of different influenza A viruses. Zoonotic transmission of Avian influenza has taken place with limited human to human transmission. There is evidence showing that the barrier of human transmissibility by some of these avian viruses is not very high, and therefore emergence into humans is possible, with most if not all of the population immunologically naïve. The humoral immune response to influenza is defined by the imprinting of the antibody response to immunodominant epitopes. Such responses can impair immunity, providing less adequate protection against seasonal and pandemic infections, as well as poorer immunity induced by seasonal vaccines. There are instances where imprinting can be advantageous and even offer protection against pandemic or avian viruses, particularly when conserved epitopes to the HA stalk are exploited. Manipulating the antibody response to recognizing conserved stalk epitopes on influenza HA is therefore a strategy being used for universal influenza vaccines. In the second Chapter of this thesis, a mosaic nanoparticle immunization strategy for inducing breadth of antibody responses against HA will be described. This strategy involves the co-display of HAs from up to eight different strains on a particle platform. Although the breadth of antibody responses elicited by immunization of these particles was limited, this work provides insight into the antigenicity of such particles, and a possible alternative to current influenza vaccines.

Approximately 100 years after the 1918-1919 flu pandemic, a deadly SARS-like coronavirus, known as SARS-CoV-2, emerged in the human population resulting in a currently ongoing pandemic. This came less than two decades after the small but deadly SARS outbreak, essentially a warning call for this class of coronaviruses. Other SARS-like coronavirus strains in bats have been identified and shown to be human tropic, though resulting in an attenuated infection. Some of these viruses can infect via hACE2 but there are others that may use an unknown receptor for entry into VERO cells as well as human cell lines. There is evidence that the major barrier to zoonosis is protease compatibility, which could be gained through recombination events or errors during replication. There-

fore, future SARS-like coronaviruses (sarbecovirus) may emerge in humans, seeding future outbreaks. The antibody response to SARS-CoV-2 is robust and protective. Furthermore, there is the presence of conserved epitopes particularly on the RBD that can be targeted by antibodies that are cross-neutralizing against many SARS-like coronaviruses. Exploiting these cross-reactive epitopes is one strategy that can be used for developing a universal coronavirus vaccine. In Chapter 3 of this thesis, a similar mosaic nanoparticle immunization strategy will be described, that attempts to elicit cross-reactive antibodies against the SARS-like coronavirus family. The mosaic nanoparticles co-display the RBDs of eight different sarbecovirus strains including SARS-CoV-2. Immunization with these mosaic-RBD nanoparticles elicited polyclonal antibody responses that were cross-reactive as well as cross-neutralizing against sarbecoviruses strains both present and not present on the particles.

## PUBLISHED CONTENT AND CONTRIBUTIONS

### **Chapter 2:**

**Cohen, A.A.**, Yang, Z., Gnanapragasam, P.N.P., Ou, S., Dam, K.-M.A., Wang, H., and Bjorkman, P.J. (2021). Construction, characterization, and immunization of nanoparticles that display a diverse array of influenza HA trimers. *Plos One* *16*, e0247963. 10.1371/journal.pone.0247963  
My contribution to this work was as the lead researcher of the project. I conceived and designed the study, prepared the reagents, analyzed the data, and wrote the paper.

### **Chapter 3:**

**Cohen, A.A.**, Gnanapragasam, P.N.P., Lee, Y.E., Hoffman, P.R., Ou, S., Kakutani, L.M., Keeffe, J.R., Wu, H.-J., Howarth, M., West, A.P., et al. (2021). Mosaic nanoparticles elicit cross-reactive immune responses to zoonotic coronaviruses in mice. *Science* *371*, 735–741.  
10.1126/science.abf6840

My contribution to this work was as the lead researcher of the project. I conceived and designed the study, prepared the reagents, analyzed the data, and wrote the paper.

### **Appendix A:**

Jette, C.A., **Cohen, A.A.**, Gnanapragasam, P.N.P., Muecksch, F., Lee, Y.E., Huey-Tubman, K.E., Schmidt, F., Hatzioannou, T., Bienasz, P.D., Nussenzweig, M.C., et al. (2021). Broad cross-reactivity across sarbecoviruses exhibited by a subset of COVID-19 donor-derived neutralizing antibodies. *Biorxiv* 2021.04.23.441195

My contribution to this work was as the co-lead researcher of the project. I helped conceive and design the cross-reactive ELISA and neutralization experiments to both sarbecovirus RBDs as well as variants of concern, prepared the reagents, and analyzed the data.

## **Appendix B:**

Wang, H., **Cohen, A.A.**, Galimidi, R.P., Gristick, H.B., Jensen, G.J., and Bjorkman, P.J. (2016). Cryo-EM structure of a CD4-bound open HIV-1 envelope trimer reveals structural rearrangements of the gp120 V1V2 loop. *Proceedings of the National Academy of Sciences of the United States of America* *113*, E7151–E7158.

I designed the experiment, prepared the reagents, and analyzed the data,

## TABLE OF CONTENTS

Acknowledgements.....	iii
Abstract .....	iv
Published Content and Contributions.....	vi
Table of Contents.....	viii
Chapter I: Introduction.....	1
Part A: Influenza.....	1
The 1918-1919 Flu Pandemic.....	1
Virus Structure and Genome.....	5
Cell Entry and Viral Replication.....	7
Epidemiology, Disease, and Animal Reservoir.....	9
Adaptation of Influenza Viruses to New Hosts.....	14
Antigenic Shift and Drift.....	17
Innate Immune Responses to Influenza Viral Infection .....	19
The Adaptive Immune Response: T-cells.....	23
The Adaptive Immune Response: B-cells and Antibodies .....	25
Immunodominance.....	33
Original Antigenic Sin and Antigenic Imprinting .....	38
Influenza Vaccines .....	47
Universal Flu Vaccines.....	49
Part B: SARS-like Coronaviruses .....	51
The SARS Outbreak .....	51
Discovery of SARS-like Coronaviruses.....	54
The COVID-19 Pandemic .....	56

Virus structure and Genome .....	59
Coronavirus life cycle .....	61
The Coronavirus Spike Protein and Its Effect on Viral Tropism, Transmission, and Pathogenesis	
.....	66
Innate Immune Response to SARS-CoV-2 .....	75
The Adaptive Immune Response: T-cells.....	78
The Adaptive Immune Response: B-cell and Antibody Responses.....	81
SARS-CoV-2 Vaccines .....	85
Bibliography.....	88
Chapter II: Construction, characterization, and immunization of nanoparticles that display a diverse array of influenza HA trimers.....	114
Abstract .....	116
Introduction.....	117
Materials and Methods.....	120
Results and Discussion.....	128
Conclusion.....	139
References .....	143
Figure Legends .....	146
Supporting Information.....	154
Chapter III: Mosaic nanoparticles elicit cross-reactive immune responses to zoonotic coronaviruses in mice .....	164
Abstract .....	166
Main Text.....	167
References and Notes .....	174
Figure Legends .....	179

Supplementary Materials .....	188
Materials and Methods .....	189
Supplementary Figures .....	195
Appendix A: Broad cross-reactivity across sarbecoviruses exhibited by a subset of COVID-19 donor-derived neutralizing antibodies .....	208
Summary .....	210
Introduction .....	211
Results .....	212
Discussion .....	219
Figure Legends .....	222
References .....	229
Methods .....	238
Supplemental Items .....	243
Appendix B: Cryo-EM structure of a CD4-bound open HIV-1 Envelope trimer reveals structural rearrangements of the gp120 V1V2 loop .....	260
Abstract .....	263
Introduction .....	264
Results .....	266
Discussion .....	272
Methods .....	275
References .....	283
Figure Legends .....	286
Supplemental Items .....	293

## Chapter 1: Introduction

### Part A: Influenza

#### The 1918-1919 Flu Pandemic:

In March 1918 approximately 100 soldiers at Camp Funston in Fort Riley became ill with flu-like symptoms, and over the next few weeks cases increased fivefold (1). Over the next half a year sporadic flu cases were reported globally in North America, Europe, and Asia. This first wave of influenza was marked by a high rate of illness with a relatively normal death rate in comparison to the seasonal influenza (1, 2). This first outbreak of a pandemic flu strain occurred in the backdrop of World War 1, where hundreds of thousands of American soldiers were being deployed overseas to Europe each month.

Starting September 1918, a second wave of influenza emerged, spreading across the northern hemisphere within the next two months. In the US, around 195,000 people died in October alone. The wave was accompanied by a shortage of trained medical personnel such as nurses, cities across the US began banning public gatherings and closing businesses as the morgues began piling up with bodies (3). Cities that employed non-pharmaceutical interventions such as social distancing sooner fared better, a good example being the longer shallow death rate in St Louis, which employed measures sooner, with respect to the sharp and rapid death rate seen in Philadelphia, which responded much later (3). Simultaneously, outbreaks occurred globally with a high death rate being reported in many countries much higher than the initial first wave of infections (3).

By winter of 1919, a third wave of infections arrived, though not quite as deadly as the second wave. Infections finally subsided by the summer of 1919. By the end of 1919 it is estimated that 500 million people (a third of the human population) had become infected, with an estimated 50-100 million deaths occurring globally (with case fatalities at around 2.5 %), making it the deadliest pandemic in modern history (3).

What was so unusual about this pandemic was that there were three waves of infections occurring within a very short one year period (3). This is in contrast to previous pandemics, such as the Russian Flu pandemic which occurred between 1889-1893, that featured waves of infections emerging yearly rather than every few months (3). Clinically, disease caused by all three waves were similar enough to conclude that they were all caused by the same infectious agent. Mild forms of disease caused by the 1918 flu pandemic were directly comparable between waves as well to mild illnesses caused by previous pandemic and inter-pandemic seasonal outbreaks (3). However it was the rapid progression from mild to severe disease, as well as the marked fatality in younger individuals, that was the main difference in the second and third wave of the 1918-1919 flu pandemic (3).

In a normal flu season, influenza infections result in uncomplicated illness that typically resolves on its own in young and healthy adults (3). Symptoms range from being asymptomatic to mild with vague symptoms, or incapacitating with almost a week of fever, chills, headache, muscle ache, and respiratory symptoms such as cough (3). In some cases for infants or elderly, seasonal influenza infection can result in severe complications such as pneumonia, which is typically a result of a bacterial secondary infection(4, 5).

The infection caused by the 1918-1919 flu pandemic was far more severe in young healthy individuals (4, 5). The majority of deaths had similar clinical pathologies, the most common being acute bronchopneumonia with necrosis of the epithelium and vascular tissue, microvasculitis, hemorrhaging, and edema. Bacterial cultures from lung samples taken by autopsy generally correlated with the pulmonary lesions, with *Streptococcus pneumoniae*, *S. pyogenes*, *Haemophilus influenzae*, and *Staphylococcus aureus* being identified as possible sources of the bacterial superinfection (4, 5). About 10-15 % of fatal cases featured a different pathology characterized by an acute respiratory distress syndrome (ARDS). In this case patients typically showed cyanosis or a blue-gray facial discoloration, and essentially drowned in their own blood which filled their lung tissue and

bronchi. It is unclear whether these cases completely lacked bacterial invasion, and were only caused by viral pathology (4, 5).

Typically during seasonal influenza, deaths with respect to age match a U shaped curve, where most deaths are in the very young or very old, with few deaths in between (4, 5). What was unprecedented, and mostly still is, is that the 1918 pandemic featured a W shaped curve, similar to the U-shaped curve, but with another peak of deaths for young adults aged around 20-40. Death rates in ages 15-34 were 20 times greater than previous years (4, 5). What's more surprising is that the majority of deaths were in the <65 age range, representing a case where the overall fatality risk was greater in the <65 age group than the >65 age group. In other flu pandemics or seasonal epidemics the fatality risk is usually reversed. The actual incidence of infection was much higher for the <35 age group, and interestingly the 5-14 age group had a low mortality even though incidence was high. Interestingly the >65 age group accounted for much lower incidence of infection, with the >75 age group showing lower case fatality rates than previous outbreaks (4, 5). This suggests some prior immunity or exposure to a similar virus in their childhood for people in the >65 age group ((4, 5)).

It wasn't until 1995 that the sequence of the virus was determined using archival autopsy specimens from 1918, showing that the virus was ancestral to the circulating swine and human H1N1 viruses which cause seasonal influenza epidemics. The Taubenberger's group in the Armed Forces Institute of Pathology successfully isolated RNA from autopsy samples that were obtained from army soldiers that died in the 1918 pandemic, as well as an Inuit woman who was preserved in permafrost in Alaska (6).

Nearly simultaneous outbreaks during March-April 1918 in North America, Europe, and Asia make it hard to pinpoint the true origin of the pandemic virus (4, 5), but a possible origin is in North America. It is also unclear whether the first, second and third waves were all caused by the same virus since all of the RNA positive samples were obtained from the second wave. It is unlikely that they were different viruses given that variation of influenza occurs over years and not months (4,

5). The animal host that acted as the source of the outbreak is also unclear since there is no sequence information from possible animal sources at or before 1918 ((4, 5). Sequence analysis suggests the virus is avian-like, however it is unlike any other virus strain obtained from mammals and birds in the past century indicated by the large number of silent nucleotide changes throughout the viral genome. This suggests the virus jumped directly to humans from an unidentified avian host (4, 5).

More recent analysis using a molecular clock approach dates the origin of the HA segment of the 1918 pandemic flu strain to before 1907. Seroarcheology also suggests that this segment appeared in the human population in the early 1900s. The rest of the viral genome appears to have arisen from the same avian source much closer to the pandemic (around 1913), suggesting that the 1918 pandemic strain is a reassortment of a human adapted strain, circulating in the early 1900s, combined with a novel avian strain which emerged shortly before the pandemic (7). Phylogenetically the 1918 pandemic strain emerged as the dominant swine flu strain after the pandemic. For humans, the predominant seasonal strains were more distantly related to the 1918 pandemic strain, especially after 1922.

Reconstructing the virus with reverse genetics allowed for the probing of why this virus was particularly so infectious, by seeing how it infects mice and NHP models. In both animal models intranasal viral challenge with reconstructed 1918 influenza resulted in a rapidly progressing pulmonary disease (4). This caused acute respiratory distress and death characterized by high viral growth which rapidly spread throughout the respiratory tract in infected animals, resulting in necrotizing bronchitis, alveolitis, and extensive edema and hemorrhaging in the lungs. These were all reported for individuals infected who succumbed to illness. Infection in NHPs elicited high expression of innate factors such as pro-inflammatory cytokines and chemokines, also known as a “cytokine storm,” but fewer viral specific Type-1 interferon responses, suggesting that dysregulated but strong innate immune responses may have been the trigger for the high fatality of the 1918 Flu Pandemic (4).

## **Virus Structure and Genome**

Influenza in humans is caused by different viruses, influenza A and influenza B. It is estimated that yearly influenza infects billions of people worldwide resulting in millions of severe illnesses and hundreds of thousands of deaths. Influenza viruses belong to the orthomyxoviridae family of viruses, and are enveloped, negative-sense single stranded RNA viruses with a segmented genome (13.6 kb total). Influenza A viruses contain eight RNA genome segments. Each segment encodes the RNA polymerase subunits (PB1, PB2, and PA), the hemagglutinin glycoprotein (HA) which facilitates viral attachment and entry, the neuraminidase glycoprotein (NA) which facilitates detachment of budding viruses, viral nucleoprotein (NP), matrix protein M1 and M2, the nonstructural protein NS1, and the nuclear export protein (NEP) (Figure 1) (8).

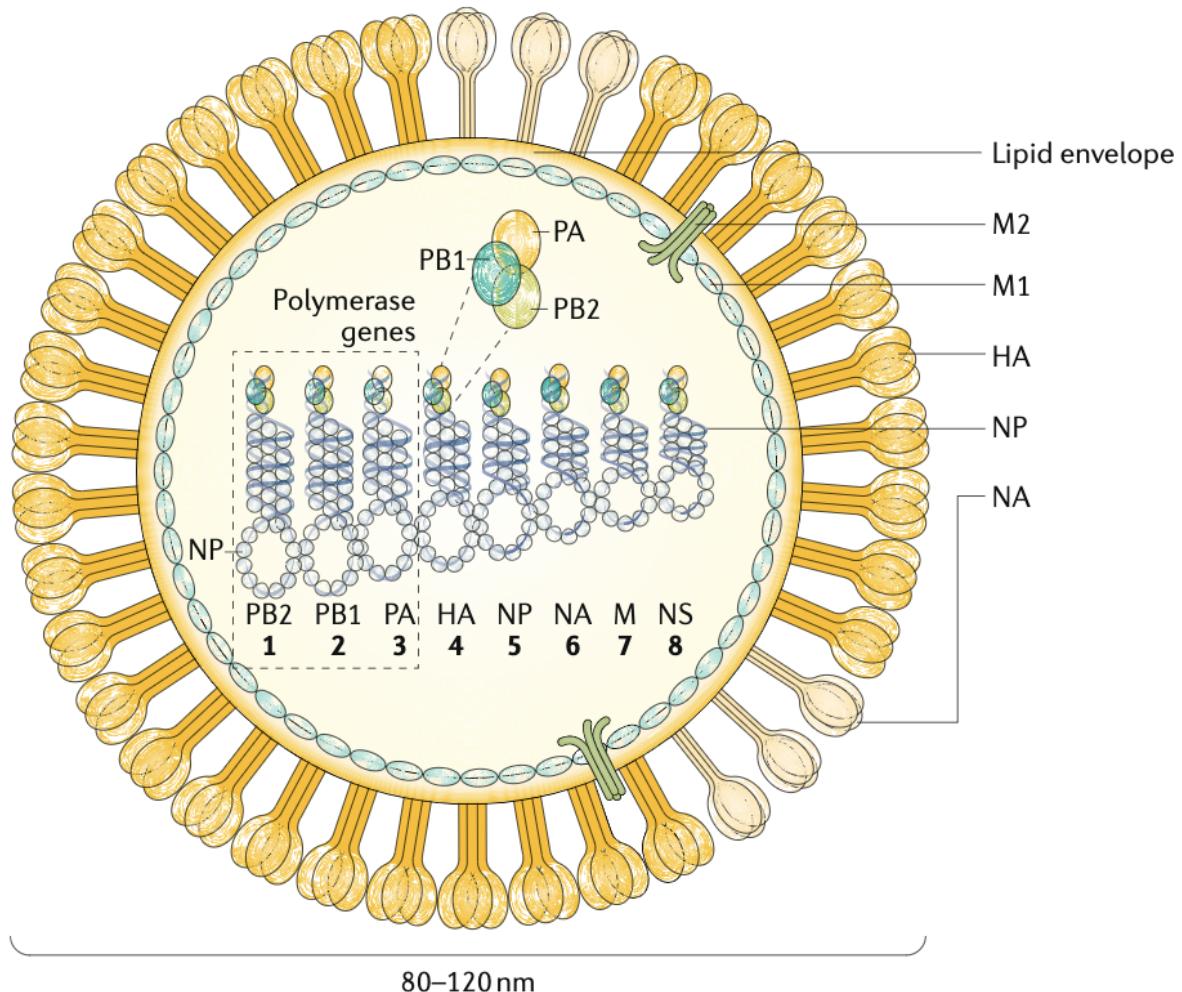


Figure 1: Schematic of the influenza Virus (8)

HA can be divided into the globular head and stalk domain. HA and NA are the most variable antigenically and are the 2 proteins located on the surface of the virion. Furthermore, the HA and NA genes are classified into different subtypes that are antigenically distinct. To date there are a total of 18 different HA subtypes and 11 different NA subtypes. Because they are the main proteins located on the surface of the virus, they are the main targets of protective antibodies elicited by the immune response. Each influenza isolate is named after its type of genus, place of isolation, isolate number, and year of isolation (8).

Influenza A viruses are able to circulate in many different animal species such as humans, pigs, migratory birds and waterfowl, poultry, domestic animals, and even sea mammals and bats. Wild migratory birds and aquatic birds are considered to be the main viral reservoir (8).

Because the Influenza genome is segmented, the virus can undergo reassortment, i.e. segments between different strains can be interchanged. This happens when two different viruses of the same type (for instance influenza A) infect the same cell. 16 of the 18 HA subtypes and nine of the 11 NA strains are able to reassort with each other. The other two HAs and NAs that are found in bats are unable to combine with the avian influenza A viruses. The animal reservoir of influenza contains an array of antigenically distinct HA and NA subtypes that can be exchanged when infecting the same host, sometimes resulting in novel pandemic strain (8). For Influenza A the HA subtypes can be further classified into two groups group 1 and group 2, which are characterized by mostly invariable stalk domains, but highly diverse head domains.

### **Cell entry and Viral Replication**

Epithelial cells in the respiratory tract (for humans) or the gut (for birds) are the main targets of influenza infection. In the first step of infection the influenza virus attaches to the target cell via binding of the HA to sialic acids on the cell surface present as either oligosaccharides or glycoproteins. For mammalian influenza viruses the  $\alpha$ 2,6 sialoside linkages are best recognized since they are most abundant in mammalian respiratory epithelial cells, particularly in the upper respiratory, with  $\alpha$ 2,3 sialoside linkages being more common in the lower respiratory tract (8, 9). In contrast, bird influenza viruses preferably recognize  $\alpha$ 2,3 sialoside linkages since they are abundant in bird respiratory and intestinal epithelial cells, with  $\alpha$ 2,6 present at lower levels in both tracts (8, 9). After viral attachment, the influenza virion is endocytosed into the cell as endosomes. Trafficking and acidification of the endosomes result in a conformational change in the HA and induces fusion between the viral envelope and the endosome. Since pH of the endosome varies between species,

pH of the endosome is another host determinant of viral tropism (8). HA protein can induce the clumping of RBCs via glycoproteins on the cell surface, a process called hemagglutination.

After viral-host fusion, the eight segments of viral ribonucleoproteins (vRNPs) are released into the cytoplasm and get imported into the nucleus where transcription and replication of the viral genome takes place, using the enzymatic activity of the viral polymerase that comes attached to the vRNPs (8). Replication occurs through the synthesis of a positive-sense intermediate, known as the complementary ribonucleoprotein complex (cRNP). In the nucleus the vRNPs are transcribed into capped and polyadenylated positive-sense mRNAs which get exported back into the cytoplasm and translated into the viral proteins. Newly synthesized PB1, PB2, and PA as well as NP are then imported back into the nucleus and amplify the replication of the viral genome (8).

The synthesized viral glycoproteins HA and NA as well as matrix proteins M1 and M2 get shuffled and inserted into the plasma membrane (8). Cleavage of the uncleaved HA (HA0) into HA1 and HA2 needs to take place in order for the HA to be functional. Therefore, the HA cleavage site is one of the main determinants of tropism, based on what host cell proteases the HA can utilize. All influenza viruses have cleave sites that are recognized by extracellular proteases found in the respiratory and gut epithelium (8).

The non-structural proteins, NS1, PB-F2, and PA-x can serve to regulate the innate and cellular responses, and dampen the host antiviral response (8). PB-F2 has been implicated in the induction of cell death .

In the final stages of viral infection, M1 and NEPs localize to the nucleus where they associate with the vRNPs, and mediate their export from the nucleus to the cytoplasm, via recycling endosomes(8). At the plasma membrane the set of eight vRNPs bundle together. Finally, budding takes place with the set of eight vRNPs being packaged into enveloped particles derived from the host plasma membrane and studded with viral transmembrane proteins, forming new virions. To prevent unproductive attachment of HA to receptors present on glycoproteins from other virions or the infected host cell, the NA cleaves off the sialic acids, allowing for efficient viral spread (8).

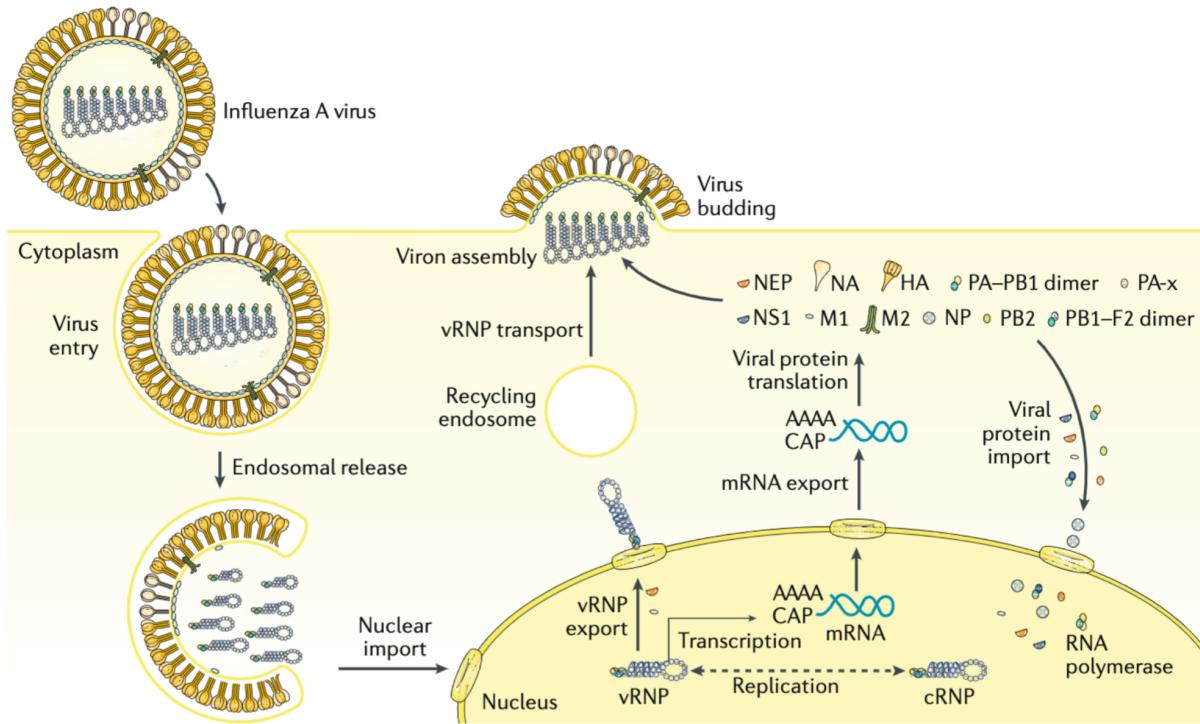


Figure 2: Viral life cycle (8)

Viral infection results in lysis and cell death, which is implicated in influenza pathology. The infection is also highly proinflammatory, resulting in the recruitment of many immune cells types, which in excess can cause immunopathology and pneumonia (8).

### Influenza Epidemiology, Disease and Animal Reservoir

Globally, influenza is estimated to infect nearly 1 billion people a year, resulting in approximately 3-5 million severe cases, with a median death around 400,000 (10). For the yearly seasonal influenza outbreaks, most infections occur in children with particularly severe infections occurring in the very young or very old. Furthermore the main drivers of transmission are children, evident

from the reduction of hospitalization and deaths in older adults when vaccination coverage in children is high (8, 11).

Influenza disease is typically characterized initially by nondescript symptoms such as fever, chills, muscle pain (myalgia), headaches, malaise (feeling unwell), and loss of appetite. Symptom onset is sudden, with respiratory symptoms appearing such as a dry cough, a sore and dry throat, and nasal discharge and obstruction. The most common respiratory symptom is coughing, and it is sometimes accompanied by a burning sensation in the sternum. In older or immunocompromised patients, symptoms can initially start out as mild but then rapidly progress to severe respiratory illness.

There are several comorbidities that increase the risk of hospitalization or mortality from influenza, such as obesity, cardiovascular disease, or immunosuppression (12), (8, 11)(12). Pregnancy has also been considered to be a risk factor for severe disease possibly due to the immunosuppression associated with pregnancies. Furthermore there is an increased risk of hospitalization and death for the elderly (>65), as well as an increased risk of hospitalization for pneumonia for the very young (<1 yo), however with a low mortality, with an average of 100 deaths in children per year (since 2010) (12), (8, 11). Young children who have had few previous exposures to influenza are very susceptible to infection, typically with high fever and long periods of viral shedding. Older individuals are also at higher risk for infection, due to associated comorbidities, and waning immunity (immunosenescence), or both.

Generally, severe disease caused by influenza can be attributed to either viral pneumonia, or pneumonia caused by secondary bacterial infections. Influenza infection increases the susceptibility to bacterial coinfection in multiple different ways: the increase of bacterial attachment via virus induced tissue damage and cleavage of sialic acids as well as dampening of immune responses (13). However bacterial coinfection may also benefit viral pathogenesis and transmission, via the secretion of bacterial proteases that may amplify activate and prime influenza virus infection (13).

Viral induced pneumonia in contrast can be characterized by high amounts of viral replication accompanied by an over-exuberant pro-inflammatory response also known as a cytokine storm (14).

Currently, H1N1 and H3N2 are the predominant strains circulating globally. From 1957 to 1968, H2N2 viruses were predominant until they were replaced by H3N2 viruses in 1968 , with H1N1 viruses reemerging around 1977 (8, 15). It is unclear whether Eastern Asia, Southeast Asia, and the tropics act as a main source of the yearly epidemics, or whether other regions of the globe act as a source such as Africa or South America, given uneven surveillance in those regions. Generally seasonal influenza outbreaks occur during the winter months, with low temperatures and humidity, conditions that favor transmission (8, 15). In temperate regions there are usually two outbreaks per year, one in the Northern Hemisphere and one in the Southern Hemisphere, both occurring during the respective winter months. However, outbreak patterns in the tropics are far more complicated and harder to predict. In general climatological conditions such as temperatures, rainfall, and maximum humidity are big factors that influence transmission patterns (8, 15).

During pandemic years, the incidence of influenza is much higher due to the lack of preexisting immunity. In the past century there have been 4 influenza pandemics: the 1918 flu pandemic (H1N1), the 1957 pandemic (H2N2), the 1968 Flu pandemic (H3N2), and finally the 2009 swine flu pandemic (H1N1). Typically the pandemic strains rapidly spread worldwide from their point of origin, and following global spread shift to a seasonal epidemic type of transmission (8, 15).

Influenza viruses have been found in a wide range of different animals including humans, pigs, horses, dogs, sea mammals, and a wide variety of domesticated birds. Wild aquatic birds such as Anseriformes (geese, ducks, and swans) as well as Charadriiformes (gulls) appear to be the main reservoir, carrying viruses with H1-H16 and N1-N9 subtypes (8, 9).

In birds there are two major types of influenza outbreaks: low pathogenic avian influenza (LPAI), and high pathogenic avian influenza (HPAI). Infection of aquatic wild birds such as waterfowl with LPAI are typically asymptomatic, are restricted to the gut and transmission to poultry such as

chickens, and also result in largely symptom free infections, however in some cases can cause substantial disease. LPAI in poultry can sometimes adapt and evolve to become HPAI resulting in systemic infection in poultry birds, with multiple organ and tissue damage, disease, and death (8, 16). Viruses from the subtype H1 to H4, H6 and H8-H16 are LPAI viruses (9). HPAI have always been from either H5Nx or H7Nx subtypes, though these subtypes are not always highly pathogenic and the vast majority of other strains from these subtypes are LPAI. One of the main factors distinguishing LPAI strains from HPAI strains is the introduction of a multibasic cleavage site in HPAIs, which allow for the cleavage and maturing of HA by ubiquitous proteases such as furin. The adaptation of HPAIs to using ubiquitous proteases allows spread outside the usual gut and respiratory epithelium to multiple different organs including the brain, which is often fatal in chickens, turkey, and domestic ducks (8, 16)). LPAIs instead have a single Arg in the HA0 cleavage site and require environmental proteases such as trypsin for cleavage. In some cases HPAI emerge in poultry almost immediately after introduction from wild birds, and in other cases the LPAI viruses have been circulating in poultry for months before acquiring high pathogenicity (8, 16)). In general, it appears that HPAI viruses are not separate lineages, but rather emerge from non-pathogenic strains as they are adapting to a new host (poultry).

Wild aquatic and migratory birds such as waterfowl and gulls are the main reservoir of diverse subtypes of influenza. Transmission of avian influenza between wild birds is complicated, but it is suggested that fecal/cloacal-oral transmission from contaminated lake or drinking water is a possible route of transmission. Domestic ducks can be an intermediary for influenza transmission between wild birds and poultry as well. Transmission from wild birds to poultry is dependent on wild bird activity such as migratory patterns and can be spread through either direct contact with wild birds, or indirect contact that is introduced via humans mechanically to poultry farms (8, 16). Transmission is accelerated through practices of raising many different domestic poultry species together either in farms or live markets and can be spread through different farms that lack biosafety measures, measures designed to protect animals and humans from infection . In some cases,

LPAIs introduced into poultry become endemic, such as the case for H9N2. HPAI outbreaks have caused a substantial disease burden and staggering economic losses throughout Eurasia, Africa, and North America, with hundreds of millions of birds killed and tens of billions of dollars in economic losses in the first decade of the 21st century (8, 16).

Zoonotic transmission of avian influenza (particularly of the H5N1 or N7N9 subtypes) have occurred and resulted in severe infections with limited human to human transmission (8). Between 2003-2020 the H5N1 HPAI virus caused 862 infections in humans with 455 deaths, and the H7N9 LPAI which causes asymptomatic disease in chickens resulted in 239 cases with 134 deaths (from 2003-2021) (17, 18).

Besides birds, pigs are another important reservoir of influenza viruses, typically of the H1N1, H3N2, and H1N2 subtypes. Influenza infection in pigs can also be severe with a similar respiratory illness as for humans characterized by high fever and pneumonia (Herfst 2012). It was noted that the 1918 H1N1 pandemic virus emerged in pigs around the same time as it did in humans, and the 2009 swine flu pandemic virus jumped from pigs to humans via aerosol, suggesting that there is direct transmission of influenza viruses between pigs and humans (Herfst 2012).

Transmission of influenza between humans is normally through the respiratory route, either through direct contact, indirect contact via fomites, or airborne transmission via large respiratory droplets ( $>5\mu$ ) or aerosols ( $<5\mu$ ) (Herfst 2012). Depending on the route of entry, the virus typically targets epithelial cells on the surface of the gut or respiratory tract. In some cases, in human influenza viruses can also infect the conjunctiva of the eye, resulting in conjunctivitis (shown in a case of human H7N9 infection). Severe influenza disease is associated with infection of the lower respiratory tract resulting in severe inflammation in the lung mediated by immune cell infiltration (Herfst 2012).

Sustained human to human transmission of HPAIs such as H5N1 have not been detected, and it is likely that the virus has not acquired the ability to transmit via respiratory droplets or aerosol (9). Although these LPAIs and HPAIs have not been responsible for the last 4 influenza pandemics,

it is still a possibility that the next influenza pandemic could emerge from an avian strain. It would be particularly concerning if an H5N1 or H7N9 strain emerged as pandemic, because the population-based immunity would be low, possibly resulting in high levels of transmission as well as morbidity and mortality (Herfst 2012).

### **Adaptation of Influenza Viruses to New Hosts**

Influenza viruses have adapted to different mammalian species as well as domestic birds, presumably jumping from wild aquatic birds. The viral RNA Polymerase facilitates this type of inter-species transmission, because of its relatively low fidelity during viral RNA replication, resulting in incorporating errors in the viral genome. The error prone nature of the viral RNA polymerase allows for high viral species diversity that can act as a mutant swarm (often referred to as viral quasi-species), which can allow for adaptations to new hosts against selection pressures. Given the nucleotide error rate estimated to be between  $1/10^3$  to  $1/10^5$  and the genome size of around 14 kb, it is unlikely that viral progeny of an infected cell is completely identical to the parent strains. Furthermore, more fit viruses can sometimes aid less fit viruses during inter species transmission, thus maintaining high diversity within the viral population that exists in a quasi-equilibrium (8, 19). This in combination with the ability of influenza viruses to reassort can allow for adaptation following zoonotic transmission to new host species.

Since the type of SA linkages determine the host tropism of influenza viruses, one of the major ways in which an avian virus would jump into humans would be to switch recognition of SA from  $\alpha 2,3$  to  $\alpha 2,6$ . The switch from recognizing  $\alpha 2,3$  to  $\alpha 2,6$  SA receptors are typically achieved via mutations in the receptor binding site (RBS) of HA: normally E190/G225/Q226/G228 for avian influenza (recognizing  $\alpha 2,3$ ), whereas for human influenza (recognizing  $\alpha 2,6$ ) either D190/D225/Q226/G228 or E190/G225/L226/S228 (20). Since  $\alpha 2,6$  linkages are abundant in the upper respiratory tract in humans and other mammals, a switch to  $\alpha 2,6$  recognition could allow for

release of virus into respiratory droplets produced in the upper airway therefore facilitating transmission between human hosts (9). Besides receptor affinity of HA, pH dependence of viral fusion and HA stability is another factor that can allow for jumping into a new host. For HA mutants that display fusion at a higher pHs (pH5.5-6) replicate better in duck than in mammals, whereas mutants that display fusion at a lower pHs (pH 5.0-5.5) replicate better in mammals (21). Finally, as described in the avian influenza section, protease cleavage of HA0 is another determinant of host range and compatibility (8).

Another factor that can greatly influence host cell range is the activity of NA, where poor NA activity in a host can allow for aggregation of viral particles which can limit transmission as is the case with H5N1 infections in humans (9). Therefore, optimal HA and NA activity can help to facilitate release of single particles, allowing for greater spread.

Besides mutations in the viral glycoproteins, amino acid changes in the polymerase subunits are also a major determinant of host tropism and transmission. Avian influenza typically replicated around 41°C (temperature in the gut) whereas for humans influenza must replicate at 33°C (temperature of the upper respiratory tract) (9). One such adaptation of the polymerase is the E627K mutation in PB2 which allows enhanced replication of H7N9 at lower temperatures (22). Furthermore, changes in NP have been associated with susceptibility to the host antiviral response. In addition, changes in M1 and M2, can change the morphology of the virus, which can be beneficial for transmission. Together all of these changes can be attained through viral reassortment as was the case with pandemic influenza viruses that emerged in 1918, 1957, 1968, and 2009 (8).

There are many inter-species interfaces which influenza viruses have crossed, such as the wild aquatic bird and domestic poultry (H5N1, H7N9, H10N8, H9N6), with swine to human (in the cause of the 2009 triple reassortment H1N1 virus, with horses to dogs (in the case of the horse endemic H3N8 virus) (8). Furthermore H17N10, H18N11 subtype influenza viruses have been isolated from fruit bats from Central and South America (23, 24). These viruses are unable to reassort

with the conventional influenza A viruses, however they are able to infect multiple different mammalian cell lines (humans and dogs) (23, 24)). Furthermore MHC2 has been identified as the putative receptor for these bat influenza viruses, with MHC2s from bats, humans, pigs, and chickens being able to be used as the virus receptor and allow for entry into non-susceptible cell lines((23, 24). This suggests that it may be possible for a bat H17N10, H18N11 virus to cross species barriers and become adapted to infecting other mammals and even birds.

In 2012 a group in Erasmus published a study showing the possibility of an H5N1 virus gaining the ability for airborne transmission in a mammalian host. They started with an authentic H5N1 strain, A/Indonesia/5/2005, isolated from a human case in Indonesia. They used ferrets as the model animal since they are susceptible to humans and avian influenza infection, and they develop respiratory disease and pathology similar to as shown in humans, and they are capable of transmitting human influenza viruses to other ferrets (9). Using targeted mutagenesis and serial transmission they wanted to see whether the H5N1 virus could gain the ability to transmit efficiently. They used both a wt H5N1 and a mutant H5N1 containing mutations in the HA (Q222L, G224S) which allow for recognition of α2,6 receptors, and PB2 (E627K) which allows for better viral replication. They serially passaged the wt and mutant H5N1 in ferrets a total of 10 times and monitored the viral titers through nasal turbinates and washes. Interestingly, the mutant H5N1 was showing greater ability to replicate in the upper respiratory tract indicated by a significant increase in viral titers in the nasal washes that were apparent after the 4th passage. For the wt H5N1 serial passage did not affect virus titers significantly (9).

In order to look at transmission between ferrets the final passage nasal wash (containing either passages wt or mutant H5N1) was used to inoculate a new set of ferrets housed with naïve ferrets. The passage 10 wt H5N1 was not able to transmit and none of the naïve ferrets got infected from the inoculated ferrets. By contrast the passage 10 mutant H5N1 was able to transmit efficiently and 75% of the naïve animals got infected. The nasal wash from these animals infected via transmission were further used to inoculate a new set of ferrets which were housed with naïve ferrets

after inoculation. Transmission occurred again with the naive ferrets getting infected by the inoculated ferrets indicating robust transmission of the passaged mutant H5N1 viruses. Sequencing of the airborne transmissible viruses from the ferrets showed only two consistent mutations in addition to the three mutations introduced, both located in the HA (H103Y and T156A), suggesting that only five substitutions might be required for H5N1 to become airborne. The T156A mutation, which removes an N-linked glycan at position N154, was shown to increase affinity to both  $\alpha$ 2,3 and  $\alpha$ 2,6 SA linkages (9). Of note none of the animals infected via airborne transmission died, and symptoms were limited to lethargy, loss of appetite and ruffled fur, in comparison to the neurological disease and death normally seen when ferrets are infected with wt H5N1. Intranasal inoculation of ferrets with a very high dose of the airborne transmissible H5N1 did result in pneumonia and death, though this is less biologically relevant. The airborne transmissible H5N1 was sensitive to both oseltamivir and antiserum from H5N1 vaccination. The airborne transmissible H5N1 was not sensitive to serum from naive individuals suggesting a lack of preexisting humoral immunity to these viruses (9).

Concerningly mutations such as the T156A and E627K have been identified via screening of H5N1 in birds suggesting that it may be simply a matter of chance before a virus like this emerges in the wild (9). It should be noted that this study only looked at one possible mammalian host as a model, and its possible these results don't directly translate into what would happen in a human host. It does however highlight the importance of doing these types of gain of function studies to identify mutations of concern in wildly circulating viruses as a method of pandemic preparedness.

### **Antigenic Drift and Antigenic Shift**

Influenza evades the humoral immune response elicited by prior infection or vaccination via antigenic drift, the process of accumulating mutation in HA and NA. This requires that the vaccine formulation be updated yearly in order to match emerging variants. Since HA is the major component of influenza vaccines, and is the main target for antibodies, changes in HA sequences are

monitored globally, and validated for immune evasion via hemagglutination inhibition assays. NA to a lesser extent is also a target of antigenic drift ((8, 16)).

A majority of the antigenic drift of influenza is focused on the HA particularly near the RBS in the head domain, with sometimes only a few amino acid substitutions being necessary for antibody evasion. For example, a single substitution that adds a glycosylation site in the head domain of an H3N2 virus circulating in the 2016-2017 season, rendered vaccination with the H3N2 egg produced virus lacking that glycosylation site, less effective due to the poor binding of vaccine induced antibodies to the circulating H3N2 that had that glycan (25). In humans, antigenic evolution is quite rapid, allowing for antigenic drift to take place in a yearly time frame possibly due to human population size, lifespan, and antibody pressures from vaccination. Interestingly in wild birds, antigenic variation of strains is not very large, however upon introduction in poultry variation is increased, possibly due to antibody pressures from vaccination in poultry.

Influenza viruses undergo far more rapid antigenic changes through the process of antigenic shift. Antigenic shift is carried out when two different strains of influenza infect the same cell in a host where reassortment of the viral gene segments occurs, with the viral progeny being a mix of the 2 parent viruses. Influenza pandemics are often associated with cases of antigenic shift, with one such example being the origin of the pandemic influenza as described earlier.

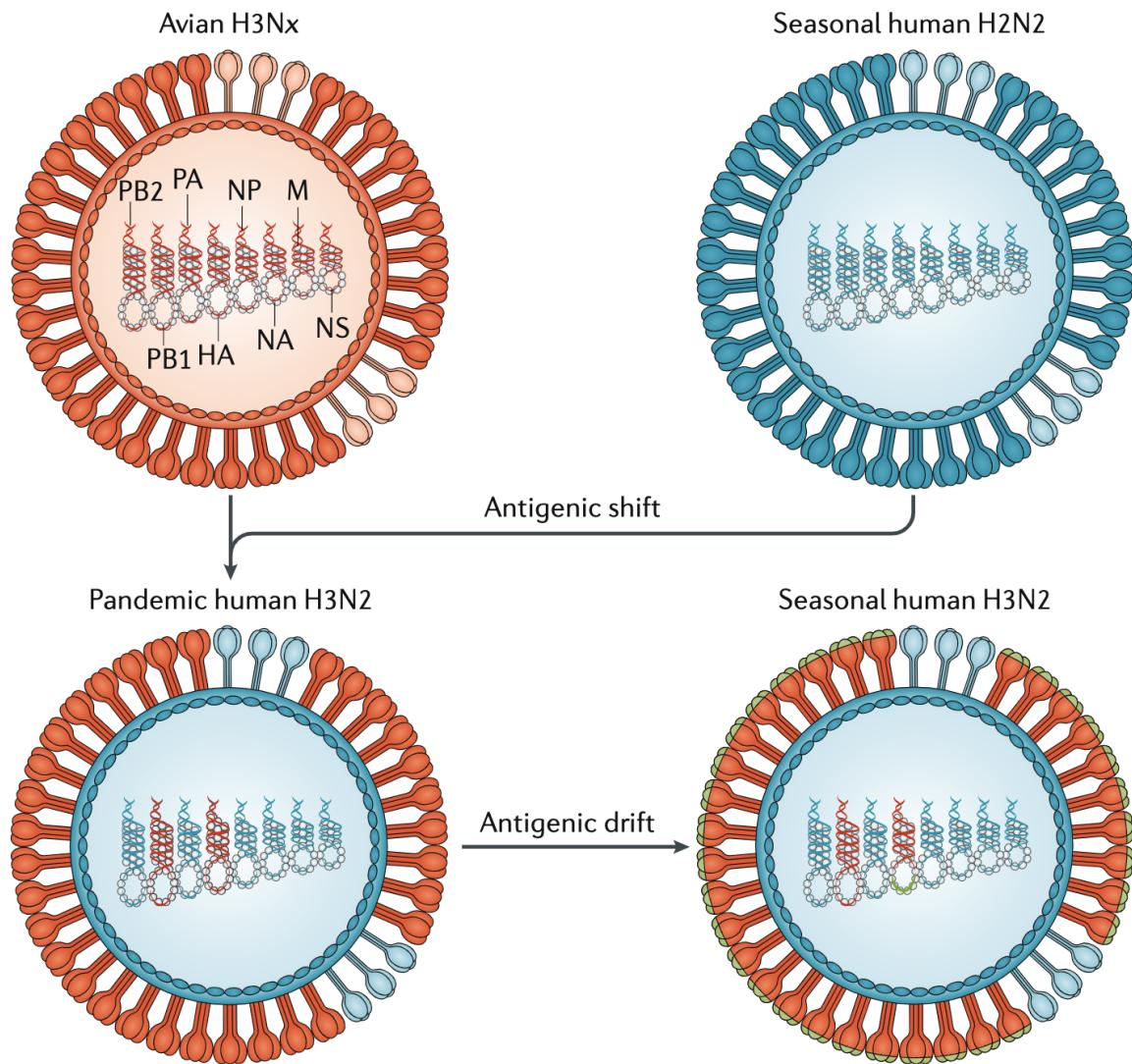


Figure 3: Schematic illustration of antigenic shift and antigenic drift ((8, 16)).

#### Innate immune responses to influenza virus infection

For influenza, Type I (IFN $\alpha$  and IFN $\beta$ ) and Type III interferons (IFN $\lambda$ ) are the main inhibitors of viral replication, and are the major interferons secreted both in-vitro and in-vivo following viral replication (8, 26). During viral infection, the main type I interferon producers are plasmacytoid dendritic cells (pDCs) and macrophages, while the main Type III interferon producers are airway epithelial cells ((8, 26)). It is thought that the Type I and Type III responses are elicited by the same signaling events intracellularly, and in the case of influenza infection upregulate virtually the same IFN-stimulated genes (ISGs). Type I interferon receptors are expressed in all cell types whereas Type III receptors are only found in the respiratory and gut epithelial cells, which possibly modulates the interferon response so that it is restricted to the site of infection (8, 26).

Interferon production is stimulated via various receptors (Pattern recognition receptors or PRRs) recognizing pathogen associated molecular patterns (PAMPs). One of the main PAMPs for RNA viruses including influenza is the viral RNA features not present in cellular RNA such as a 5' Triphosphate moiety or dsRNA. Some of the PRRs recognizing these PAMPs include the Toll-like receptor family, and the cytosolic RNA sensors. The TLRs activated by influenza infection include TLR7, which recognizes ssRNA and is important for producing high levels of Type I interferons in pDCs, and TLR3 which recognizes dsRNA and mediates the proinflammatory response.

The RNA sensors include the retinoic acid-inducible gene or RIG-I, and the melanoma differentiation-associated protein 5 or MDA5 and are crucial for the interferon responses in non-dendritic cells such as the airway epithelium. RIG-I and MDA5 recognize the 5' triphosphate moiety and dsRNA formed by the 5' and 3' ends of the influenza viral RNAs, which is normally required for viral RNA replication (8, 26). RIG-I and MDA5 are very similar, containing two caspase activation and recruitment domains (CARDs), a C-terminal regulatory domain (CTD) and a helicase domain (26). Binding of the helicase domain and the CTD to RNA ligands exposes the CARD domains. Modulation of the CARDs through post translational modifications (dephosphorylation and ubiquitination) activates its function, for example through the polyubiquitination of Lys63 via the E3 ubiquitin ligases Trim25 and RNF135, which activates RIG-I (26). Downstream the activated CARD domain of RIG-

I and MDA5 associate with the mitochondrial antiviral signaling complex (MAVS), inducing the oligomerization of MAVS, as well as activation of large complexes responsible for activation of the interferon regulatory factors (IFRs) and NF- $\kappa$ B, which then activate transcription of type I and type III IFNs as well as numerous interferon stimulated genes (ISGs) (26). The RNA sensors are crucial to the overall antiviral response against influenza with RIG-I being predominantly used, though there may be redundancy with MDA5.

Secreted Type 1 IFNs signal via IFN receptors (IFNAR) switching on janus kinase (JAK1) and tyrosine kinase (Tyk2) activating STAT1, STAT2, and IRF9, forming a complex that translocates into the nucleus to activate transcription at interferon-stimulated response elements (ISREs). As a result hundreds of ISGs get turned on with antiviral functions (27). Activation of ISGs can lead to further amplification loops as well (27). The IFN response is heterogenous and can depend on cell type, microenvironments, etc., with different amplitudes and kinetics. The ability to fine tune the IFN response is critical in disease outcomes for viral infections because there are negative consequences for both the overactivation and underactivation of the IFN response (27) .

The innate immune response is a crucial barrier to overcoming infection by viruses, and as such influenza has multiple proteins whose role is to evade these responses. NS1 is an RNA binding protein that prevents recognition of viral RNA by RIG-I, the pattern recognition receptor responsible for activation of type-1 interferons. NS1 inhibits the RIG-I mediated signaling cascade by binding and inactivating the E3 ubiquitin ligases, trim25 and RNF135, required for RIG-I activation after viral RNA binding. This shuts down the downstream activation of interferon and interferon inducible genes that have antiviral activity. Examples are the Mx family of GTPases that inhibit the viral nucleocapsid mediated nuclear entry and replication, protein kinase R (PKR) which ultimately shuts down translation preventing viral protein synthesis, IFN-induced transmembrane members (IFTIM) which interferes with viral-host fusion, among others. NS1 also shuts down host mRNA synthesis further preventing the synthesis of interferons and interferon stimulated genes (8, 26). NS1 may

also target the JAK-STAT pathway (janus kinase and signal transducer and activation of transcription) as well as the antiviral activity of interferon stimulated genes such as PKR or RNaseL.

Other influenza viral proteins besides NS1 can affect the innate immune response. PB1-F2 a non-structural protein encoded by an alternative reading frame of PB1, as well as PB2 both suppress MAVS. Another reading frame of PA known as PA-x is another viral factor that shuts off mRNA expression, and therefore interferon and interferon stimulated gene expression, via its RNA endonuclease activity.

Although epithelial cells are the main targets of infection both tissue resident macrophages and dendritic cells can be infected to varying degrees, which has implications for the immune response driven by these innate immune cell types, as well as pathogenesis. For instance, in macrophages replication of seasonal IAVs are typically abortive due to various host restriction factors, however productive replication in macrophages such as H5N1 HPAI strains may result in overproduction of proinflammatory cytokines or hypercytokinemia and a hyperinflammatory response (28). Infection of dendritic cells can be important for antigen presentation via MHC-1, though the less understood more controversial phenomenon of cross-presentation is also very important, and does not require DC infection (29). IAV infection may also impair the ability of antigen presentation of MHCs, via either direct presentation or cross presentation which can diminish the cytotoxic T-cell (CD8+ T-cells) response (29).

Exposure of macrophages and dendritic cells to IAV infection results in secretion of proinflammatory cytokines. Proinflammatory cytokines in turn recruit natural killer cells (NK-cells), monocytes as well as neutrophils to the lung tissue where they kill infected cells and promote viral clearance. There is a delicate balance between the function of these activated innate immune cells in clearing viral infection and resulting in a hyperinflammatory response. Overactivation of these cell types can result in an over-exuberant immune response (often regarded as a cytokine storm). The cytokines responsible for this are largely IFNs, Tumor necrosis factors (TNFs), Interleukins (ILs), and chemokines (30). In patients the cytokine storm promotes high infiltration of innate immune

cell types resulting in immunopathology and severe disease such as acute lung injury (ALI) (8, 14). As an example high neutrophil counts in the lungs of patients with severe H7N9 and H1N1 infection correlated with lung damage, mediated by high levels of neutrophil extracellular traps (NETs) which increased permeability across the airway epithelial layer (31).

Proinflammatory cytokines can have various roles in the progression of disease. For example, IL-1 is released early in infection and IL-6 is released later, where IL-1 is protective and IL-6 null have more severe disease upon IAV challenge. TNF- $\alpha$  on the other hand is responsible though not required for escalation of the cytokine storm (30). Abnormally high levels of chemokines such as IL8 and MCP-1, which recruit neutrophils and monocytes, respectively, seem to be associated with more severe pathology where levels (30). Altogether it is the circuit of cytokines/chemokines working together which help to mediate the cytokine storm, with the full circuit not fully elucidated (30).

The cytokine storm does not occur often in the seasonal influenza infection. On the other hand it is likely that the cytokine storm is responsible for the high death rate in the 1918 influenza pandemic (5) as well as the high mortality with H5N1 and H7N9 HPAI infection in humans (30). Polymorphisms in PRRs may actually be responsible for some of the variability in why cytokine storms only happen in some infected individuals but not others (30). As such targeting the cytokine storm with various immunomodulatory agents such as corticosteroids and sphingosine-1-phosphate receptor 1 agonists has become a treatment option for severe influenza disease (30).

### **The adaptive immune response: T-cells**

For IAV infection and vaccination CD4+ T-helper cells provide help to both the humoral B-cell response as well as the cellular CD8+ killer T-cell response (8). The main contribution that CD4+ T-cells provide to overall immunity and protection against influenza is through helping the B-cell and antibody response. The help is mediated by the interaction between the T-cell receptor (TCR) on CD4+ T-cells and the cognate peptide antigen that is presented on MHC-2 by influenza

specific B-cells (32). This interaction drives class switching for both short lived antibody secreting plasma cells as well as longer lived affinity-matured bone marrow plasma cells and memory B-cells derived from germinal centers (32).

CD4+ T-cells also play an important role in establishing CD8+ T-cell immunity, by enhancing recruitment of CD8+ T-cells to germinal centers, and by promoting the priming and expansion of CD8+ T-cells, establishing CD8+ T-cell long-lived memory, and aiding the positioning of CD8+ T-cells in the respiratory tract (32). Help given by CD4+ T-cells is further important for establishing the cytotoxic activity of CD8+ T-cells as well as allowing them to home in and extravasate to their virally infected target cells (32).

CD4+ T-cells can also be involved in recruiting innate effector cells into the lung, modulating their effector function for optimal viral clearance such as IFN- $\gamma$  production (32). Finally CD4+ T-cells can also have effector functions, and play a role in killing virally infected cells in the lung (32).

CD4 T-cells are activated by antigen presenting cells or APCs displaying optimally 13-25 amino acid long peptides presented on major histocompatibility complex class-2 or MHC-2. The CD4+ T-cell response to IAV infection is highly diverse and is directed towards most if not all of the viral proteins including HA, NA, NP, M1, M2, and the viral polymerase (32). However the breadth, as well as the particular epitopes recognized on each viral antigen, depends on the MHC-2 haplotypes, as well as the complex history of infection and vaccination the individual was exposed to in their lifetime (32). Given the heterogeneity of the MHC-2 alleles in humans with 8-10 being inherited, there are hundreds of possible epitopes for CD4+ T-cells in influenza. Furthermore activated CD4+ T-cells of wide viral epitope specificity travel and stay in the lung where they can elicit effector functions to virally infected cells, as well as encounter antigen presenting cells or APCs (32). Secondary responses to reinfection and vaccination match the degree of diversity elicited in primary CD4+ T-cell responses, suggesting that CD4 T-cell memory is strong and long lived (32).

CD8+ T-cells or Cytotoxic T-cell lymphocytes (CTLs) are required for effective viral clearance, although without CD8 T-cells antibody mediated effector functions can compensate, albeit

more slowly. CD8 T-cells recognize viral 8-12 aa long peptides that are present on the major histocompatibility complex class-1, or MHC-1, present on all cell types. The recognition occurs via the T-cell receptor, or TCR, which is complementary to the peptide MHC-1 complex. The peptides are derived from components of the viral proteins expressed in infected cells. The CD8 T-cell response to influenza can be directed to a diverse set of antigens, though NP, M1, and the viral polymerase seem to be the most prominent. Peptide antigens recognized by CD8+ T-cells can be from viral proteins that are more conserved across various IAV subtypes and therefore less subject to antigenic drift. There is evidence that CTLs elicited against conserved epitopes can result in better outcomes upon H7N9 influenza infection, whereas lack of these CD8+ T-cell responses is associated with higher disease severity (33).

Since CTLs can be directed towards conserved epitopes and are associated with protection against heterosubtypic IAV infections, they are an attractive target to elicit for a broadly protective universal flu vaccine. Live attenuated influenza virus vaccination can elicit these types of CTLs however they are not long lived and do not contribute to memory. One of the strategies that will be needed is to improve the CD8 T-cell memory pool so that cross-reactive CTLs can be recalled, resulting in more rapid viral clearance, and milder disease. Therefore, strategies at improving CTL memory, which is poorly understood, are required.

### **The Adaptive Immune Response: B-cells and antibodies**

During influenza infection IAV specific naive B-cells or preexisting memory B-cells are typically activated by recognition of viral antigens through the B-cell receptor (BCR or Ig receptor) together with helping interactions with cognate CD4 T-cells (34). During the primary response some of these activated B-cells rapidly differentiate into short-lived plasmablasts (also called antibody secreting cells or ASCs) that secrete IgM initially and later IgG and IgA after class switching. Other activated B-cells migrate to germinal centers in secondary lymphoid tissue and undergo germinal

center reaction and affinity maturation (34). The product of the germinal center reactions are B-cells that secrete higher affinity antibodies and comprise the secondary response which peaks around day 14. The first wave of antibodies found in the serum comes from the rapidly activated plasmablasts and peaks approximately 7 days, especially if the antibodies originate from memory B-cells. The role of this plasmablast driven primary response is to rapidly increase the presence of protective serum antibodies (34). Later subsets of plasmablasts that are class switched and affinity matured typically come from the germinal centers. A small subset of these plasmablasts can migrate to the bone marrow where they become long-lived plasma cells or bone marrow plasma cells (BMPCs). BMPCs secrete antibodies for long periods and maintain a long-term steady-state serum antibody level that can be protective against subsequent reinfection and disease (34). Another fate of activated B-cells is the differentiation to long-lived memory B-cells (MBCs) that do not secrete antibody but participate in immune surveillance and remain at the periphery. Reactivation of MBCs occurs rapidly after exposure to antigen, with differentiation into plasmablasts that secrete high-affinity antibodies as well as more memory B-cells (34). For the case of IAV, the memory B-cell compartment is highly diverse with B-cells recognizing many different antigens and epitopes, with a high prevalence of broadly-reactive antibodies, whereas the serum antibody repertoire is more restricted strain specific. This might allow for the ability to rapidly target conserved epitope on antigenically drifted or shifted strains amounting to some protection (34).

There are 5 isotypes of human antibodies: IgM, IgG, IgA, IgD, and IgE. IgM IgG and IgA are the main protective antibodies against influenza. IgG is Y shaped, existing as a monomer with two fragments of antigen binding domains (fabs) as its arms and an Fc as its tail (34). IgG is the main antibody in the serum (75%) and there are 4 subtypes: IgG1 (67% of IgG), IgG2 (22% of IgG), IgG3 (7% of IgG), and IgG4 (4% of IgG) (34). A majority of the IgG response of influenza is IgG1 mediates with a lower amount of IgG3, and negligible amounts of IgG2 and IgG4. IgG1 has a long serum half-life (around 21 days) and interacts strongly with Fc receptors (FcRs) hence it can have both neutralizing activity and Fc effector functions (34). IgG3 has an even stronger affinity to FcRs, due to

its flexible hinge, but it has a shorter half-life (around 7 days), and its effector function is mediated by a glycosylation at position N297 (34).

IgM can exist as a monomer, pentamer or hexamer and accounts for 10% of antibodies in the serum. IgM is typically secreted in the early response to influenza infection, and typically has lower affinity than IgG antibodies (34). However IgM can cross the mucosal barrier and be a strong activator of complement, especially in its multimeric state, and therefore IgM is important for immunity against influenza (34). IgA represents 15% of the serum antibodies and is divided into 2 subtypes IgA1 and IgA2 (34). IgA1 exists as a monomer in the serum whereas IgA2 is secreted to the mucosa as dimers, and also to a smaller extent larger multimers. Since IgAs are highly glycosylated their binding to HA might be further enhanced through these glycans. IgA also interacts with myeloid cells such as monocytes through FcRa (34).

Since the mucosal surface of the respiratory tract is the main point of entry, protection along this barrier is important. In the lower respiratory tract IgG is dominant and therefore most protection is mediated by IgG which crosses the mucosa through neonatal Fc receptors (FcR). Antibodies found in the respiratory tract tend to differ from those found in the serum (34). In the upper respiratory tract IgA (particularly secretory IgA1) is the main protective antibody since its levels are dominant over IgG (34). IgA is secreted by plasmablasts that are located in the lamina propria of the mucosa-associated lymphoid tissue (MALT), and crosses the mucosal surface via polymeric Ig receptor (pIgR) (34). Mucosal immunity is not very well studied, though IgA targeting might be similar to IgG targeting though possibly more broad owing to the avidity effects of dimeric IgA binding (34).

B-cell responses are crucial for immunity elicited by IAV infection and vaccines. Antibodies secreted by activated B-cells are protective against infection, and passive transfer of antibodies to a naive host is protective against viral challenge (34). The main antigens detected by B-cells are HA and NA, since both of these antigens are present on the surface of the virion as well as the surface of infected cells (large numbers of HA trimers and NA tetramers that form a dense array of virions). Portions of the M2 ion channel are also exposed on the surface of the virus though they

are less targeted. Internal proteins such as NP, M1, polymerase (PB1, PB2, PA), and NS1 are accessible usually after cells die after infection, and can be recognized by B-cells as well (34).

Based on the viral life cycle not all the viral antigens are expressed at similar levels or presented equally, having dramatic consequences to their recognition by the antibody response.

A majority of the antibody response is elicited towards the HA protein during natural infection (34), with a lower amount of antibodies elicited against the NA as well as the internal viral proteins. Infection with IAV typically induces the seroconversion of antibodies particularly to the HA subtype (34). This is usually detected via enzyme linked immunosorbent assay or ELISA, and more informatively hemagglutinin inhibition assays (HI) (34).

The influenza HA is composed of two structurally distinct domains, the globular HA head composed of the central part of HA1 and the elongated stalk composed of the HA2 subunit as well as the N-term and C-term portions of the HA1 (34). The antibodies targeting the head portion of HA are highly neutralizing typically by blocking receptor recognition by the RBS. Therefore head targeting antibodies typically prevent viral attachment to host cells or the first step of the viral life cycle (34). The activity of these antibodies is often assayed by their ability to inhibit hemagglutination of HA in a Hemagglutination Inhibition (HI) assay (34). Hemagglutination inhibition is the process in which antibodies block the ability of IAV HAs bind to sialic acids on red blood cells (RBCs) that causes RBC clumping. In vitro neutralization, or the ability to block infection of an authentic virus or pseudovirus from entering cells, correlates well with hemagglutination inhibition (34). HI and neutralization potency of antibodies elicited by vaccination or neutralization both correlate well with protection in animal models as well as in people (34). However antibodies against the head can also be neutralizing but without displaying HI activity (34). Therefore neutralization and HI are surrogates of protection when determining the efficacy of influenza vaccines, and these are typically mediated by head targeting antibodies (34). For example a 1:40 HI titer is considered to reduce the infection rate of seasonal IAV by 50% (34). Head targeting antibodies with HI and neutralizing activity are the

main correlates of protection used to determine vaccine efficacy for seasonal influenza vaccines (34).

Five neutralizing and non-overlapping antigenic sites or epitopes on the globular head domain of HA have been identified in classical studies using the H1N1 influenza virus PR8 (A/Puerto Rico/8/1934) (35). Sa and Sb sites are located at the tip head region of the HA homotrimer, whereas Ca1, Ca2, and Cb are located towards the stem portion of the head that attaches it to the stalk (34, 35). For H3 HAs the main epitopes on the head domain are A, B, C, D, and E. The head domain of HA tends to be highly variable between subtypes, as well as being prone to antigenic drift, therefore, antibodies that target this site are highly strain-specific (34). Addition of N-linked glycans, through a process called glycan shielding, is a common way in which head epitopes are mutated to avoid antibody responses that target them (34).

There is also a class of head domain targeting antibodies that recognize the receptor binding site (RBS) the region where the hemagglutinin binds its sialic acid receptor, and therefore these antibodies are highly potent. RBS targeting antibodies are broad in binding, being able to cross-react to both group 1 and group 2 viruses since this epitope is relatively conserved (36). These RBS targeting antibodies use long CDR3s to mimic the binding of sialic acid in the sialic acid binding pocket (36). They also appear to be commonly elicited in the human memory B-cell repertoire, being able to arise from many different germlines (37). Targeting the RBS epitope may be an attractive target for a universal coronavirus vaccine.

A third class of HA antibodies are elicited against the stalk domain. Since this portion of the HA is highly conserved between subtypes and even across groups, antibodies that target it are highly broad in reactivity. In general HA stalk antibodies are usually restricted to binding only within group 1 HAs (H1, H2, H5, H6, H8, H9, H11, H12, H13, H16, H17, and H18) or Group 2 HAs (H3, H4, H7, H10, H14, and H15) (38). There have been some stalk-specific antibodies that bind within group 1, and group 2, and even influenza B (39). Importantly stalk-reactive antibodies have been shown to

be broadly neutralizing in vitro and protective in animal models, even showing protection across different subtypes (34).

Stalk antibodies appear to also be an independent correlate of protection against IAV infection (40). Although they are not accepted by the regulatory agencies as a surrogate for influenza vaccine protection. Stalk antibody mediated protection can be due to neutralization. Neutralization of stalk antibodies is at a different stage of the viral life cycle and does not involve blocking attachment (34). One of the neutralization mechanisms is the locking of the HA in the prefusion conformation, which prevents fusion of the viral and endosomal membrane and subsequent release of the viral genome (34). Another mechanism for stalk antibody mediated neutralization is the inhibition of viral egress or the inhibition of viral release from infected cells, however the reason behind this is unclear. There is some evidence that inhibition of viral egress may be due to steric blocking of the NA enzymatic site, which is important for viral detachment from infected cells (34). A third mechanism is the blocking of the HA1-HA2 cleavage site, preventing the HA0 from being cleaved rendering the virus non-infectious since cleavage is necessary for the conformational changes that guide viral fusion (34).

The mechanisms for protection with these broadly protective anti-stalk antibodies, however, is not only due to neutralization (blocking the virus from binding or entry into cells) (41). These other mechanisms of cross-reactive antibody mediated protection include antibody dependent cell-mediated cytotoxicity or ADCC (the killing of infected cells by effector cells such as NK cells through surface bound antibody), antibody dependent cellular phagocytosis or ADCP (the phagocytosis of infected cells through surface bound antibody), and complement-dependent cytotoxicity CDC (lysis of infected cells mediated by complement through surface bound antibody) (34). These mechanisms are mediated by anti-stalk antibody Fc effector function. HI positive head targeting antibodies have minimal Fc effector function, apparently because they block sialic acid binding by HA to effector cells, which is apparently required in addition to Fc FcR interactions (34). Other non-HI negative head targeting antibodies have been shown to protect via FcR functions (34).

Some anti-stalk antibodies show high in-vitro ADCC potential, as opposed to strain specific highly potent anti-head antibodies (41). Fc<sub>Y</sub>R knockout mice are not protected by the cross-reactive stalk targeting antibodies, nor are they protected from cross-reactive stalk binding antibodies that don't bind Fc<sub>Y</sub>Rs, even though they retain their in-vitro neutralizing capacity (41). This is in contrast to protection by anti-head specific antibodies which protect mice regardless of whether they can bind Fc<sub>Y</sub>Rs or not (41).

The human memory B-cell repertoire has been analyzed for the presence of Stalk-reactive antibodies using various methods, but recently carried out using stalk only based probes that can assay to binding of only stalk epitopes (42). B-cells in the memory B-cell compartment that bound these probes were common and found in 84% of the individuals screened reacting to at least one probe (42). Serum binding to these probes was common, as well as being able to cross-react to both H1N1 and H5N1, though the serum was not neutralizing in all cases with only 33% of samples able to cross-neutralize H5N1 and H1N1 (42). Furthermore older individuals appeared to have greater quantities of stalk specific antibodies and B-cells (42). Altogether stalk reactive antibodies appear to be fairly commonly elicited in most individuals from normal influenza exposure. Furthermore stalk antibodies can be elicited in certain contexts of infection (43). For example during the 2009 H1N1 pandemic anti-stalk antibodies with cross-group 1 breadth were elicited in many individuals presumably because common epitopes shared between the pandemic strain and the prior seasonal H1N1 strains were restricted to the stalk (43). There have been active strategies to redirect the response to stalk epitopes which will be discussed later (44–48).

The NA is also targeted by the antibody response, although to a lesser extent than HA. Antibodies against NA can have antiviral properties and can be neutralizing by blocking NA enzymatic activity. NA targeting antibodies are usually characterized by ELISA or Neuraminidase inhibition assays, assays designed to measure the activity of antibodies in blocking sialidase activity of NA (34). The mechanisms of blocking are either binding to the enzymatic active site or steric hindrance of the sialic acid binding. Blocking enzymatic activity of NA can have several outcomes. It

can make the influenza virus more sensitive to mucins (the sialic acid containing proteins that can trap influenza viruses on the mucosal surface) (34). Another consequence is the prevention of viral detachment from host sialylated proteins on infected cells that is mediated by NA cleavage activity. Furthermore NA is important for preventing aggregation of influenza viruses mediated via sialic acids on adjacent virions, and blocking neuraminidase activity can result in aggregation of influenza viruses(34). NA targeting antibodies can also be active through Fc effector function and have been shown to be capable of ADCC (34). NA targeting antibodies are also determined to be an independent correlate of protection as well (40).

NA tends to be less immunogenic or immunosubdominant when presented with HA; this may be due to the lower copies of NA that are presented to the HA (34). Antibodies targeting NA are however commonly found in most individuals with elderly individuals having the highest titers. Levels of NA reactive B-cell plasmablasts elicited by infection of either H1N1 or H3N2 viruses are around 14-35 %, lower than the HA-specific response (34). Antibodies targeting NA tend to be broad in specificity and inhibitory to a given subtype, and are sometimes broadly protective within a given subtype against challenge in animal models (34). Furthermore NA antibodies induced by infection can last for decades possibly due to either backboosting during subsequent infections or generation of long-lived plasmablasts (34).

The internal viral proteins can also be targets of antibodies especially during infection where infected cells can express and release them, though these viral proteins are recognized to varying degrees (34). For example antibodies targeting the NP are present after natural infection in around 80% of individuals, with durability of these antibodies lasting for at least a year (34). Furthermore the NP-specific B-cells in the memory B-cell repertoire of healthy individuals is highly diversified probably owing to repeated IAV and therefore NP exposure, which stimulates NP B-cells (34). Antibodies against M1 appear to show up in varying degrees in individuals with IAV exposure, whereas titers against M2 appear to be common but short-lived, with higher titers linked to old age (34). M2

antibodies have been shown to be effective in mice through Fc effector functions (34). Finally, antibody titers against PB1, PB2, PA, PB1-F2, NS1, and NEP have been detected although the magnitude of these responses is not well defined. Furthermore, the function of these antibodies has not been well defined. Of note, these antigens are highly conserved within IAVs and therefore antibodies targeting them are highly broad (34). Cross-reactive antibodies that are not against HA or NA can be protective in vivo without being neutralized in vitro, suggesting that mechanisms other than blocking viral attachment or entry are used (38).

Antibody responses to influenza infection are generally robust and long lived. For example, individuals infected with the 1918 pandemic influenza strain have antibodies specific to the pandemic H1N1 HA that were highly neutralizing as well as circulating B-cells nearly 90 years later (49). This offered protection to older individuals born before 1950, who were presumably exposed to similar viruses, against the 2009 H1N1 flu pandemic, whereas younger individuals who were exposed to antigenically distinct H1N1 strains were not as well protected (50). In contrast the antibody response to influenza vaccination, specifically the inactivated influenza vaccine, is much more short-lived. As an example, within a year after vaccination bone-marrow plasma B-cells (BMPCs) specific to the vaccine strain decline and return to pre vaccination levels, whereas levels of non-vaccine strain specific BMPCs were stable (51). The reasons behind the disparaging immunity induced by infection versus vaccination are unclear (34). Antibody responses against the HA and NA antigens are complex and ultimately depend on prior exposure in the individual. Since most young children (<2) have been exposed to the influenza virus, therefore re-exposure in the future elicits recall responses as well as de-novo naive responses. This can lead to the phenomenon of original antigenic sin (OAS) or imprinting.

## Immunodominance

Immunodominance refers to the different immunogenicities (i.e. preference for the immune system to target) for different antigens as well as different epitopes on the same antigens. B-cell immunodominance is not completely understood, although overcoming it is needed for a universal vaccine against influenza and HIV, as the antibody response to these viruses tends to be elicited against variable and strain-specific immunodominant epitopes (35). Two main examples for influenza are that HA is immunodominant over NA (34), and that the HA head domain is immunodominant over the HA stem. Interestingly these ID hierarchies have also been observed for lampreys which have VLRs that are derived from TLR, with HA being immunodominant over NA, and the HA head being immunodominant over the HA stalk (52).

Five neutralizing and non-overlapping antigenic sites or epitopes on the globular head domain of HA have been identified in classical studies using the H1N1 influenza virus PR8 (A/Puerto Rico/8/1934) (35). Sa and Sb sites are located at the tip head region of the HA homotrimer, whereas Ca1, Ca2, and Cb are located towards the stem portion of the head that attaches it to the stalk (34, 35).

The immunodominance hierarchy for B-cell epitopes on influenza was determined using a panel of mutant five mutant PR8 HAs that abrogate the antigenicity of the 5 dominant epitopes (35). Influenza virus infection results in a consistent hierarchy elicited against these five dominant epitopes. Influenza virus infection as well as different routes of immunization with inactivated virus all had different ID hierarchies for the five different epitopes, suggesting that context of the antibody response influences the hierarchy of epitopes being recognized (35). Different mice strains featured different ID hierarchies as well after viral challenge. Finally, blocking one antigenic site with either a monoclonal or polyclonal serum from an immunization (with a virus that only has the targeted epitope intact) does not shift the hierarchy of the other epitopes, but completely suppresses the response against the blocked epitope (35). Interestingly in this study there were no appreciable responses to the HA stalk (35) .

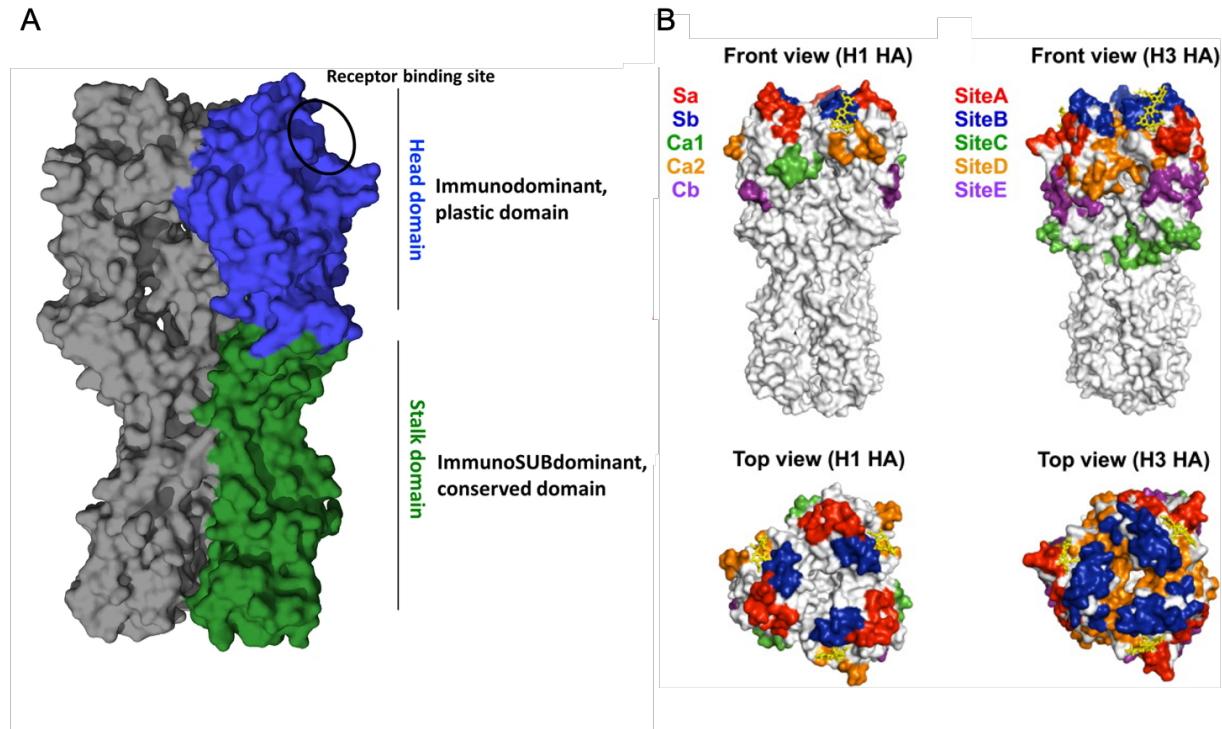


Figure 3: The immunodominance of hemagglutinin domains and epitopes. A) HA homotrimer domains are split into head and stem, adapted from (53). B) Five immunodominant antigenic sites on H1 and H3 HAs, adapted from (54)

The results of this study were corroborated in where HI inhibition titers to the mutant panel after infection were characterized in other animals models (55). In this study guinea pigs, ferrets and mice all featured a different ID hierarchy (55). Furthermore human serum pre and post vaccination was also characterized with the mutant panel where the ID hierarchy was different from other animal models (55). Interestingly the HI ID hierarchy was virtually the same pre and post vaccination, and the majority of the HI response in humans pre and post vaccination was again to the five head epitopes (55). Together these two studies showed that for humans and mice the Sa and Sb antigenic site is immunodominant (35)(55)(35). The ID hierarchy for H3 HA has also been investigated, with the antigenic site B being the most targeted (56).

In a recent study the immunodominance hierarchy was modeled using MD simulations based on the BCR affinity and avidity of various epitopes on HA presented on either a virus, a ferritin nanoparticle as an 8-mer, a soluble trimer, and a stalk only HA on a nanoparticle (57). Influenza virus particles favored head responses since they are readily accessible for bivalent binding by a BCR, over the stem which is buried. HA trimers show lower amounts of epitopes capable of being accessible through bivalent binding vs HA multimerized on a nanoparticle, whereas stalk-NPs have predominantly bivalent interactions with conserved stem epitopes (57). Furthermore ID hierarchy with stem vs head responses matched *in silico* immunizations, which in turn matched *in vivo* immunization of humanized mice (57). Priming with either HA-*np* or a divergent virus strain followed by two subsequent boosts with stem-NPs redirected the HA response to the stem, subverting the immunodominance hierarchy (57).

B-cells recognizing highly conserved epitopes on the HA stalk are present in the memory B-cell compartment, however they are overshadowed by the more immunodominant but less conserved head specific B-cells. Accordingly boosting with highly divergent HAs may preferentially expand the stalk specific B-cells since they target better conserved epitopes (58). Frequent boosting with H3N2 strains during seasonal influenza vaccination tends to rarely elicit these stalk reactive B-cells, suggesting they are immuno-subdominant (58). Furthermore, although these boosting with a highly divergent strain may elicit stalk reactive antibodies, subsequent boosting may result in out-competing of more strain specific head reactive B-cells (58).

The head of the hemagglutinin for the pandemic 2009 strain vs. previous strains is around 68% conserved, whereas the stalk is around 88% conserved, with some epitopes on the head being more conserved between strains. Based on the age of each cohort, vaccination of individuals with the 2009 pandemic H1N1 strain elicits either a broadly reactive stalk-biased response or a more strain specific head targeting response (58). The stalk-biased group were from a younger age group and were presumably only exposed to H1N1 viruses antigenically distinct from the 2009 pandemic H1N1 with more antigenically different HA heads, and therefore had lower pre-vaccinated serum

against older more antigenically distinct H1N1 strains as well as the 2009 pandemic strain (58). The head-biased cohort were from an older age group, and had higher preexisting serum antibody against the antigenically distinct H1N1 strains as well as the 2009 pandemic strain (58). Interestingly revaccination of both cohorts with the 2009 H1N1 pandemic strain biased the response against the HA head, suggesting that once there is some preexisting memory against the head it is preferentially activated over stalk specific memory (58). In conclusion, a stalk-specific broadly neutralizing response is elicited only when exposure occurs with an antigenically divergent strain that shares few epitopes on the HA head, however reboosting with the same strain will bias the response back to the head.

There are several hypotheses as to why the broadly-reactive and broadly neutralizing stalk-specific antibodies are so hard to elicit. One reason is that these epitopes are harder to reach on the virion than the better exposed head epitopes, and this has been confirmed with the lower apparent affinity of stalk reactive antibodies to whole virus vs recombinantly expressed hemagglutinin (58). A second reason is that the stalk reactive antibodies display greater polyreactivity, i.e. they can bind to off-target epitopes, than head-specific antibodies (58), (59). The polyreactive quality of these antibodies gives them the ability to bind autoantigens, and therefore B-cells with this specificity would be eliminated through B-cell tolerance (58), (59). A third possible explanation is that stalk specific antibodies are restricted in humans to certain antibody lineages such as the VH1-69 germline lineage (58).

The HA head is one of the major sites for antigenic drift. This is due to the fact that the HA head is susceptible and tolerant to mutating, and as a result highly variable (34). Therefore, antibodies that target the HA head are highly sensitive to antibody evasion. In the case of seasonal influenza if there is antigenic mismatch between the circulating strain and the vaccine strain in a given year, the vaccine effectiveness can be decreased. This is also the case for pandemic influenza strains that have undergone antigenic shift, where the antibody-based immunity in individuals

is not matched to the pandemic strain, allowing for its rapid spread. Altogether this shows the importance of the antibody response in controlling and protecting against influenza infection (34).

Exposure of humans with pre-existing seasonal flu immunity to the avian influenza strains such as H5N1, H7N9, H6N1, and H10N8 has happened. Since H5N1 and H7N9 have pandemic potential as discussed, vaccines have been prepared to target them (34). Humans lack immunity to the head domains of H5 and H7 since they are largely different from the H1 or H3 heads humans would normally be exposed to (34). The stalk domain of H5 has epitopes conserved with H1 and H2 (group 1) whereas the stalk domain of H7 has shared conserved epitopes with H3 (group 2) (34). Exposure to the H5 HA induces a recall response to the stalk-specific MBCs, resulting in an expansion of stalk-specific plasmablasts and production of broadly cross-reactive anti-stalk antibodies (60). Similarly H7N9 infection elicited similar cross-reactive anti-stalk antibodies in this case being cross reactive and cross-neutralizing to both group 1 and group 2 HAs (61). H5 immunization tends to elicit anti-stalk antibodies that are limited to pan-group 1 breadth that is restricted to the VH1-69 germline lineage (34). Exposure to H7 HAs on the other hand tends to elicit stalk specific antibodies that feature pan group 1 and group 2 breadth that came from a diverse array of germlines (VH6-2, VH1-18, and VH3-53) suggesting that group 2 HAs might be better immunogens for a universal flu vaccine (62). The reason for the difference in breadth elicited H5 vs H7 immunogens is not clear. It does appear that the titers elicited by H5 vaccination are higher than for H7 vaccination, and this may be due to higher baseline activity for H5 than for H7. Interestingly if a H5 or H7 immunogen is vaccinated a second time, the immunodominance for the head is restored, and antibodies targeting the head will be preferentially activated. This is not necessarily detrimental for the anti-stalk responses elicited in the second H5 or H7 exposure.

### **Original Antigenic Sin and Antigenic Imprinting**

Original antigenic sin (OAS) postulates that a person's first exposure to influenza takes antigenic seniority in the B-cell memory repertoire (63). In other words the first exposure to influenza virus leaves an "imprint" where subsequent exposures to newly circulating strains of influenza will result in the reactivation or recall of memory B-cells that are specific to epitopes shared with the priming strain (63). This can both be beneficial or detrimental depending on the context, and therefore imprinting may be more apt terminology for this phenomenon (63). An example of how this is beneficial is how during the 2009 pandemic influenza outbreak, the older cohort that were presumably exposed to H1N1 strains before 1950 had neutralizing antibodies against the 2009 H1N1 pandemic strain, and were better protected and less susceptible (64). Exposure of mice to these pre-1950 H1N1 strains offered protection against the 2009 H1N1 pandemic strain as well (64).

There are multiple different models of OAS. In the first model for classical OAS, the secondary response to new viruses is dominated by antibodies with low affinity to the new strains but high affinity to the primary strain(34). The low affinity response to the new strains can mature and get increased affinity towards the new strains, or conversely there may be cases where the antibodies have equal affinity to both sets of strains (34). The second model (known as back-boosting ) occurs when infection or vaccination with a new strain boosts antibody responses against previous historic strains as shown in antibody landscapes for H3N2 viruses (65)(34), (65). Antibody responses to viruses encountered earlier in life (antigenically senior strains) increase over time, due to imprinting, and following many subsequent exposures over time though responses to the new virus are not necessarily impaired (65)(34), (65). A common OAS is known as head specific imprinting, and can occur when a head epitope is shared between the primary strain and new strain (34). The primary response to the first infection induces a strong response to the head epitope which is then recalled and strongly boosted during exposure to the new strain (34). This type of imprinting can also take place within group 1 strains. An example of this is how imprinting with either H1N1 or H2N2 reduced the morbidity and mortality of an infection of H5N1 later in life. This may be due to group 1 reactive stalk specific antibodies that are found in the memory B-cell compartment. This phenomenon is also

seen with Group 2 strains, with H3N2 primary exposure being protective against morbidity and mortality caused by H7N9(66). This phenomenon is known as group imprinting. A final type of OAS is known as antigen masking, where vaccine effectiveness is reduced with consequence vaccinations of the same strain (34). This is possibly due to the presence of strain specific antibodies that can block epitopes from being recognized by B-cells, though this needs more investigation (34).

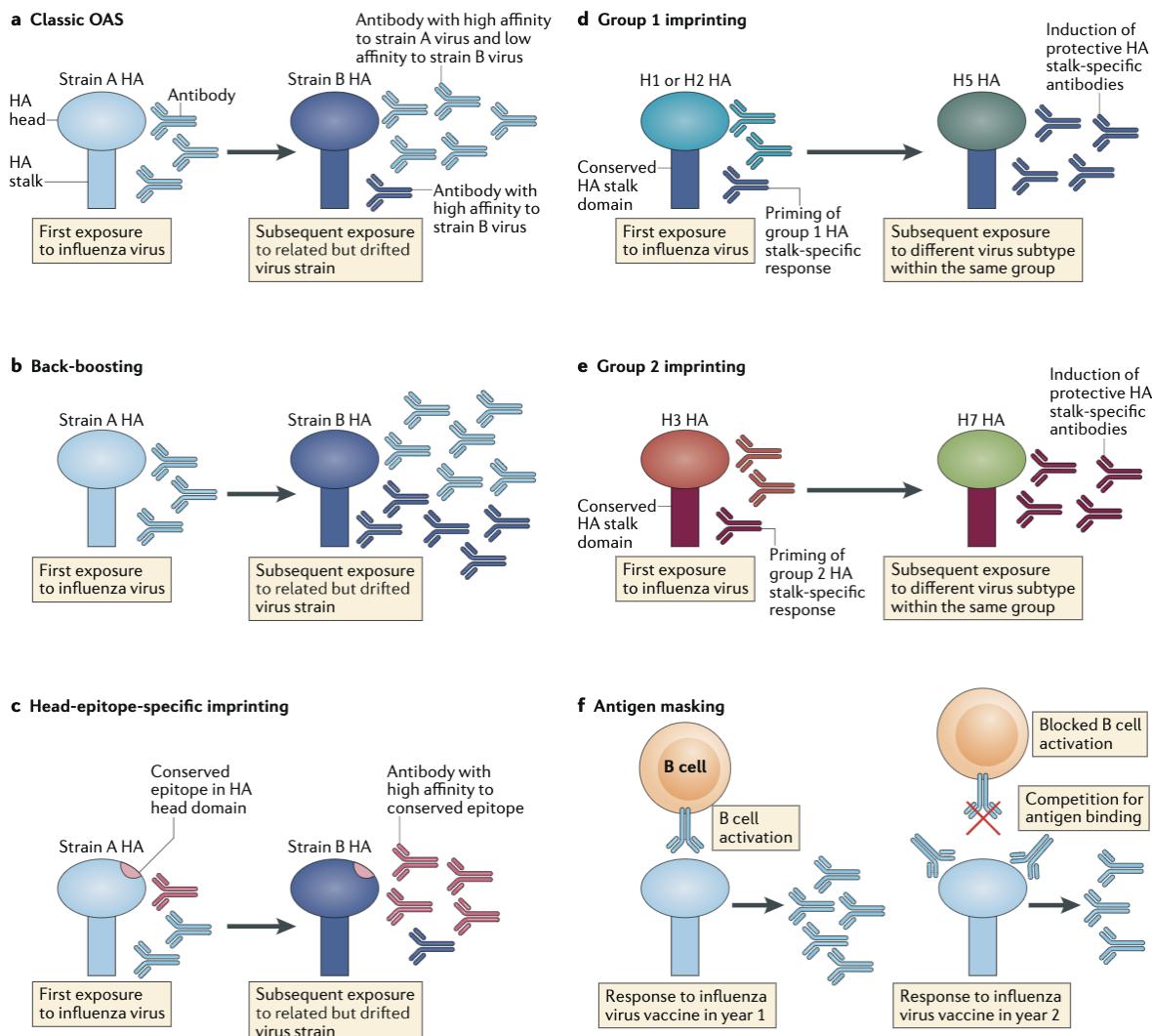


Fig-

ure 4: The different models of antigenic imprinting taken from (34).

The presence of antibodies in three different age groups, young (18-20), middle-aged (33-44), and experienced (49-64) matched the best with the strains circulating during childhood of each age cohort (38). Antibodies against H1 as well H2 along with the more related H5 were particularly high in the experience group given that the likely exposure for this group would be an H1N1 or H2N2 virus (38). Anti-H3 responses were high in the middle-aged group with lower responses against H1, which this cohort was presumably exposed to later in life, matching the circulation of only H3N2 during the childhood this age cohort. Finally the young cohort has high levels of antibodies against H1 and H3 given that H1N1 and H3N2 have been cocirculating in the past couple of decades (38). Together this provides strong evidence for imprinting or original antigenic sin where an individual is predisposed to making antibody responses biased towards the primary infecting strain. The distribution of seasonal influenza cases in different age groups can be predicted based on the primary infecting subtype during the childhood of each cohort.

The antibody response to HA depends profoundly on the exposure history of the individual, with children generally having narrower responses than older adults due to less numbers of exposures (38). Adults depending on which strains they were first exposed to and their subsequent exposure history, can have a diverse range of antibody responses to HA, that can be particularly broad (38). In some cases antibody responses to the historic strain dominate over the responses to the currently infecting strain which can sometimes be an impairment to developing a protective antibody response, a process known as original antigenic sin (38). Breadth of the antibody response also depends on which particular strain a person is exposed to (38).

In a paper published by the Krammer group the so called “cross-reactome” was assessed to determine the breadth of antibody responses from infections in animal models, infections in humans, and age stratified cohorts with presumably diverse exposure histories (38). Animal models

showed varying degrees in the induction of cross-reactive antibodies when challenged with one or two subsequent infections with divergent virus from the same subtype (two subsequent but divergent H1N1 infections, or two subsequent and divergent H3 infections) (38). Guinea pigs showed the greatest breadths followed by mice followed by ferrets, with cross-reactivity being restricted mostly to HAs within the same group (H1N1 infection results in pan-group 1 breadth and H3N2 infections result in pan-group 2 breadth) (38). Remarkably in humans clinically-confirmed infections with pandemic 2009 H1N1 in humans induced much greater HA breadth than seasonal H3N2 infections which were much narrower, showing that the induction of cross-reactivity can be strain dependent (38). In agreement, a previous study showed that the antibodies elicited by pandemic 2009 H1N1 infections featured broad reactivity to multiple H1N1 and H5N1 strains (67). Antibodies elicited in these patients were extensively somatically hypermutated (suggesting they are boosted from memory), targeted epitopes on the stalk, and were neutralizing and protective in mouse models against heterotypic challenge (67). Cross-reactive stalk-specific antibodies may be important in reducing morbidity and mortality of antigenically drifted and shifted pandemic strains (41). They are also attractive targets for a broadly protective universal influenza vaccine (41).

There are major differences within the antibody response elicited by natural infection, versus antibody responses elicited by vaccination (63). In a comparison to a cohort infected with pandemic 2015-2016 H1N1 strains or the 2014-2015 H3N2 strains and a cohort vaccinated with the quadrivalent influenza vaccine, stark differences in epitope specificity were seen (63). For both H1N1 and H3N2 infections the antibody response is biased to conserved epitopes such as those on the NP, NA, and HA stalk, with the H1N1 infections eliciting better responses against the HA head (63). These antibodies were poorly neutralizing, broadly-reactive in some cases both within and between subtypes, but were minimally protective in mice against lethal challenge (63). In contrast vaccinated individuals elicited an antibody response that predominantly against HI+ head epitopes, with minimal but largely neutralizing stalk antibodies being elicited as well (63). The vaccine induced HA antibodies were potently neutralizing, broadly reactive to HAs from heterotypic strains within the

same subtype, and protective in mice against lethal challenge (63)). Of note is that the antibodies elicited by infection appeared to recognize past strains with equal or higher affinity, and also contained more somatic hypermutations than antibodies elicited by vaccination, suggesting that the infection induced antibodies were elicited from memory B-cells from past infections that were recalled (63)). Altogether this study highlights the importance of memory B-cell recall in OAS where memory B-cells against conserved non-neutralizing epitopes outcompete more potent but strain specific B-cells during an infection (63)). Vaccination though does seem to overcome this mechanism, however the responses can be much shorter-lived (51).

The 1918 H1N1 Flu pandemic mortality rates are a particularly good example of how priming during childhood with a particular influenza subtype governs the immune responses elicited when challenged with a different strain (7). Matching the unique mortality rates for the 1918 flu pandemic to seroarcheological data is highly suggestive of OAS playing a role in outcomes (7). The >75 aged cohort were likely exposed to an H1N1 similar to the 1918 influenza H1N1, which was the likely source of the 1830 pandemic. Therefore the >75 age group would have had considerable protection against the 1918 pandemic H1N1, which explains the lower mortality rate in this group (7). Those born around 1890-1893 were most likely primed with an H3N8 pandemic strain and would have had minimal protection against the 1918 pandemic strain. This is a possible explanation as to why the peak mortality in the young adult age group belonged those aged 25-29 at 1918 (7). Those born before the 1890 pandemic would have had intermediate protection as the likely seasonal influenza strain was a H1N8 subtype which would have offered some protection, but if their first exposure was the H3N8 1890 pandemic strain, then they would lack protection (7). Supporting this, presence of H3 antibodies correlated with mortality rates from the 1918 pandemic, suggesting that primary exposure to a mismatched influenza strain greatly affects the disease outcome(7). Therefore, a possible explanation as to why the young adult age group featured such a high mortality is because they were primed with an H3N8 virus, and therefore elicited an ineffective recall response to the H1N1 pandemic virus (7).

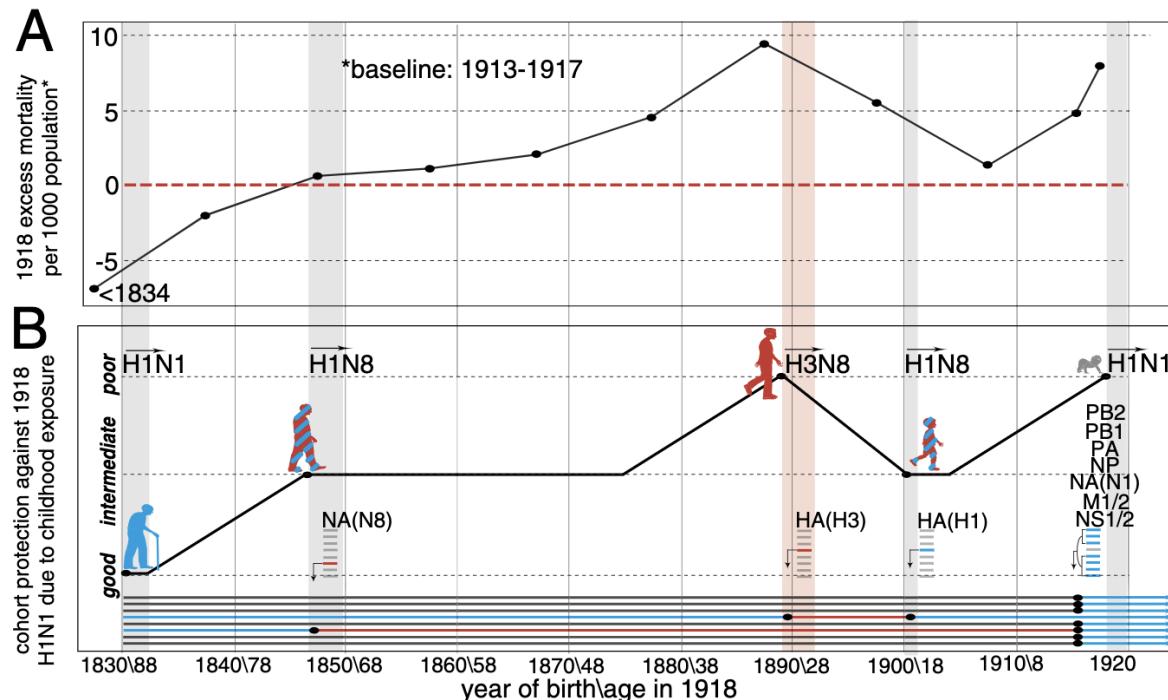


Figure 5: During the 1918 Flu Pandemic childhood exposure may have explained the unique mortality rates.

The manifestation of immunodominance as well as original antigenic sin may be elucidated by studying clonal dynamics that take place in the germinal center after infection or vaccination (68). In a groundbreaking study the B-cells were tracked in and out of germinal centers and from a primary PR8 infection followed by a boost with H1 HA (68). Infection results in a large expansion of clones in germinal centers, however during the boost almost none of the MBCs from primary infection made it back to the secondary GCs, with the B-cells in this compartment almost entirely made up of naive B-cells (68). The few MBCs that do reengage in the secondary germinal center also tend to come from a very restricted set of clones, but are substantially diversified and expanded (68). This could explain why conserved epitopes are poorly targeted on HA as continuous exposure

to influenza may prevent cross reactive B-cells from actively being engaged in secondary GC reactions (68). On the other hand the plasmablast/ and therefore serum antibody response from the boosts were completely derived from the MBCs elicited by the primary infection, and restricted to a few clones (68). The MBCs that were reengaged in either GC or PB responses seem to derive from germlines with higher overall affinity than those MBCs that were not re-engaged, suggesting that germline affinity is a strong predictor of whether a clone would be recalled (68). The limited clonality that is accessible upon boosting and is expanded, as well as the requirement for high affinity from the start, may help explain why OAS and immunodominance are so troublesome for influenza (68).

CD4+ T-cells can also be important in this phenomenon (32, 69). For example, in mice heterotypic challenge of H3N2 virus followed by H1N1 virus showed poor antibody responses against the HA of the 2nd challenge H1N1 strain, but robust antibody responses against the NP. T-cell Epitopes on the HA between the two strains are not well conserved whereas epitopes on other viral proteins such as the NP are highly conserved. CD4+ T-cells recognizing non-conserved epitopes on the HA of the heterotypic H1N1 challenge strain were not activated, in contrast to the CD4+ T-cells recognizing conserved NP epitopes. These NP+ CD4+ T-cells appeared to be recalled and were prefferentially activated suggesting that T-cells recognizing conserved T-cell epitopes out-compete naive T-cells recognizing non-conserved epitopes (69). Therefore, the impaired antibody responses against the heterotypic HA could be the result of poor T-cell help because CD4 T-cells recognize less conserved epitopes on HA and are poorly elicited (69). Furthermore it was shown that preexisting NP-specific CD4+ T-cells only stimulate B-cell responses against NP but not HA and vice versa, and therefore T-cell helper functions play a role in establishing immunodominance of B-cell epitopes (32). These have been shown in both mice as well as previously infected or vaccinated humans.

There may be a couple of reasons to explain the selectivity in CD4 T-cell help to antibody responses. One possible explanation is that dendritic cells that migrate to the germinal centers release minimal virus in the lymph nodes, and therefore there are few intact virions for B-cells to

uptake (55). Furthermore, uptake of infectious viruses by antigen specific B-cells via the BCR results in death of those B-cells after 18 hours, eliminating them from the pool of B-cells that secrete antibody or become memory B-cells (55). The major source of antigen for B-cells in influenza infection therefore appears to be membrane fragments containing viral proteins, or viral proteins released from dying cells rather than intact virions (55). This suggests that HA specific B-cells will only be presenting HA specific T-cell antigens, therefore not activated by CD4+ T-cells recognizing conserved epitopes on other viral antigens. Therefore the most efficient help of the CD4+ T-cell response for the neutralizing antibody response is when the CD4+ T-cell antigen specificity for HA matches the B-cell antigen specificity for HA, where CD4+ T-cells (55).

Another example of this is how inactivated H5N1 vaccination in humans is relatively poor at eliciting anti-H5 HA responses, with low amounts of CD4+ T-cells recognizing epitopes in H5-HA (70). This may be due to mismatch of the CD4+ T-cell memory responses against the H1N1 HA (closest related influenza virus in humans) that would not overlap as much in specificity with the H5 HA (70). There was strong boosting of CD4+ T-cells against the H5N1 NP which is far more conserved between the two strains. However a second vaccination with different clade H5N1 HA that would presumably recall H5 HA responses did expand H5 HA specific CD4+ T-cells better than vaccination of naive individuals (70). Furthermore this increase in H5 HA specific CD4+ T-cells correlated with higher neutralization titers against H5N1 virus (70). Finally NP specific T-cell responses did not correlate with H5N1 neutralizing responses (32). This suggests that improving the HA specific CD4+ T-cell response may be crucial in developing neutralization breadth. It should be noted that cross reactive epitopes on HA exist for CD4+ T-cells and can contribute to immunity against a heterotypic challenge. For example there are H3 HA specific CD4+ T-cells from seasonal influenza infection or vaccinated that are cross-reactive to H7 HA (71). Presumably these cross-reactive T-cells can be expanded by H7 boosting and can improve neutralizing antibody responses directed towards the H7 HA head (71). The amount of these T-cells that recognize cross reactive HA epitopes are bound to vary from individual to individual.

## Influenza Vaccines

The original influenza virus consisted of crudely purified and chemically inactivated virions known as whole inactivated virus vaccines or WIVs (65)(34)). For safety concerns of WIVs mainly due to the reactogenicity, split virus, subunit (inactivated influenza vaccines or IIVs) and recombinant influenza hemagglutinin (HA)-based insect cell expressed vaccines (Flublok) are the vaccines being used in North America and Europe. WIVs are still being used and manufactured in low- and middle-income countries due to the easier downstream processing steps required for these vaccines. WIVs have been used extensively in humans and are very well studied in animal models because they are easy to produce (65)(34). Depending on how they are inactivated they may retain HA hemagglutination and fusion activity as well as NA neuraminidase activity. They also contain viral RNA, making them self-adjuvanting by giving them the ability to activate RIG-I as well as TLR3,7 and 8 (65)(34). WIVs tend to elicit a balanced immune response with seroprotection rates approaching 85%. It has been shown that split virus inactivated vaccines (IIVs) do approach similar levels of protection that WIVs do.

Inactivated split virus or subunit vaccines typically administered intramuscularly (i.m) without adjuvant. They contain 15 ug of HA per strain and are given as a trivalent (H1N1, H3N2, and B) or a Quadrivalent (H1N1, H3N2, B/Yamagata, and B/Victoria lineages) (65)(34)). In some cases, for the elderly adjuvanted inactivated vaccines (with MF59, a squalene-based oil-in-water emulsion) are given. A majority of IIVs are manufactured in embryonated chicken eggs, with some inactivated vaccines (Flucelvax) produced in cell culture. They are usually produced using whole inactivated virus that is then treated with detergent and selectively purified. Depending on how they are purified, the final preparation either contains HA and NA (split virus) or just almost pure HA (subunit). The viral RNA is completely removed, which results in both decreased antigenicity but also decreased reactogenicity (65)(34)). Not much is known about the structural integrity of the HA and NA after the

purification steps for a split or subunit vaccine, and whether the crucial epitopes on the HA and NA are intact (65)(34)). Typically, vaccinations with split or subunit vaccines are biased against the HA with HI seroconversion rates (defined as 4-fold increase in HI titers) from between 20-77% and neutralization seroconversion between 47-90%. Antibody responses to NA are also variable with seroconversion rates between 20-73% (defined as a 2 fold increase in antibody titers) (65)(34)). Anti-NP titers are also detected in around 20-87% of individuals. Furthermore around 80-90% of B-cells induced by vaccination are HA specific with 1-2% being NA specific. Therefore the majority of the immune response elicited by split or inactivated vaccines are to HA (65)(34)). Inactivated influenza vaccines can elicit broadly reactive antibodies against HAs from historic virus strains in adults with preexisting immunity (65) (65)(34)). However, the responses tend to be relatively narrow and strain-specific than antibody responses induced by infections, and antibodies that target the conserved HA stalk are rarely elicited. Furthermore antigenic mismatch between the circulating virus and the vaccine strain can greatly reduce vaccine effectiveness in a given year (65)(34).

The recombinant HA vaccines are also administered intramuscularly (i.m) without adjuvant and given at higher doses with 45 ug of HA per strain, given as either a trivalent (H1N1, H3N2, and B) or a Quadrivalent formulation (H1N1, H3N2, B/Yamagata, and B/Victoria lineages) (65)(34). In some cases, in elderly the Flublok recombinant HA vaccines are given at higher doses. The recombinant HA vaccines are produced in insect cell culture and therefore contain smaller glycans than those produced in chicken eggs. Since they are HA only the immune responses elicited are only to HA, though they are comparable with the inactivated influenza vaccines (65)(34). There is some evidence showing that recombinant HA vaccines elicit broader antibody responses with better protection in the elderly. An added benefit is that fewer mismatches occur since it is directly expressed rather than being adapted to chicken eggs where it can gain mutations(65)(34)).

Live-attenuated influenza vaccines (LAIVs) used in Russia and the USA. They are typically temperature-sensitive and cold adapted meaning they can grow well at lower temperatures but poorly in higher temperatures (65)(34). For this reason, they can only infect the upper respiratory

tract (lower temperature) and not the lungs (higher temperature) and are administered as a nasal spray (Flumist) (65)(34). LAIVs are also manufactured in embryonated chicken eggs and can be given as a trivalent or quadrivalent formulation. Since the IIV and Recombinant HA vaccines are administered intramuscularly they may not induce sufficient mucosal immunity, though it is possible that they may boost mucosal immunity from previous infection (65)(34). Since LAIVs are administered through the nose and replicate in the upper respiratory tract they can provide mucosal immunity and the immunity to LAIVs is not entirely serum antibody dependent. For this reason, LAIV efficacy is determined by protection rather than correlates of protection such as HI and neutralization. Seroconversion in adults for LAIVs was low at 3-7% for HI, 3-13% for neutralization, and 0-17% for NA inhibition (65)(34)). By contrast, seroconversion rates for children were much higher for HI, neutralization as well as NA inhibition, with children even showing an induction of cross-reactive antibodies against the HA stalk, although these were not at high levels. LAIV mediated immunity is more durable than vaccination with IIVs, though it may be shorter than natural infection (65)(34)).

Vaccines against H7N9 and H5N1 have also been developed with the vaccine platforms used above and are currently in preclinical or clinical trials. Other vaccine platforms also include  $\Delta$ NS1 vaccines that have the deleted NS1 which attenuated the innate immune response. DNA vaccines, mRNA vaccines, virus like particle (VLP) vaccines, and recombinant protein vaccines as well as other adjuvanted vaccines are all being tested in clinical trials as well.

### **Universal Influenza Vaccines.**

There are ongoing major efforts to develop a universal influenza vaccine, one that would be broadly protective against many influenza strains. Such a vaccine would remove the need to revaccinate yearly, improve accessibility of the vaccine to developing countries, and greatly improve pandemic preparedness. As discussed throughout this introduction several antibody targets exist to

achieve this: the HA stalk as well as conserved head domains on the HA head, the NA, and the M2 ectodomain (65)(34)).

There are two major vaccine candidates against the HA stalk: the stalk only vaccines, and the chimeric HA vaccines that refocus the immune response towards the conserved HA stalk through sequential immunizations. In the stalk only approach, since the head is removed the immune response is redirected towards the stalk epitopes, and immunization with these immunogens elicit a broad-reactive antibody response that is also protective against heterotypic challenge (45, 48). However these stalk based immunogens had a relatively low immunogenicity in comparison to a full length HA which may be due to lower quantities of Tfh cells elicited by HA stalk immunization (72).

The chimeric HA approach involves sequential immunization with chimeric HAs that have HA heads from different subtypes but identical HA stalks (44, 73). Since the heads of the sequential immunogens are drastically different, but the stalk is conserved and identical, the idea is that the immune response is redirected towards the stalk upon sequential boosts (44, 73). Immunization of ferrets with a set of three sequential different chimeric HAs, as either a LAIV or IIV induced cross reactive antibody response that targets the stalk (44, 73). Furthermore sequential chimeric HA immunization was protective against both homotypic and heterotypic challenges (44, 73). Furthermore this chimeric HA strategy was shown to be safe and immunogenic, and elicit a broadly reactive immune response against conserved stalk epitopes in humans (46). The serum antibodies elicited with the chimeric HA immunization strategy were protective in mice against heterotypic challenge (46).

Another strategy for eliciting cross-reactive antibodies against HA is a mosaic nanoparticle approach (74, 75). In this approach a protein-based nanoparticle is used to co-display HA immunogens from different subtypes or strains on the same particle. The rationale behind this is to preferentially activate B-cells recognizing conserved epitopes on the HA over less conserved variable epitopes since B-cells recognizing conserved epitopes will be able to bind with avidity to only the

conserved epitopes (74). As such immunization with a mosaic particle displaying 8 different receptor binding domains or heads from divergent H1N1 strains featured pan-H1N1 neutralizing breadth that extended even to H5 (74). This breadth was significantly different from the breadth elicited by an admixture of homotypic particles displaying the same eight strains. Therefore this strategy seems to subvert immunodominance toward conserved elements (74). Immunization with a quadrivalent mosaic particle that featured full HAs from H1N1, H3N2, B/Yamagata, and B/Victoria strains featured breadth and potency against a panel of group 1 and group 2 HAs (74, 75). Mosaic HA nanoparticle immunization furthermore was protective against homotypic and heterotypic challenges (74, 75). However Mosaic-HA nanoparticles were as good as an admixture of homotypic particles of the same strains and inducing breadth, suggesting a different mechanism for breadth (74, 75).

Another approach is the computationally optimized broadly reactive antigen or COBRA approach. This strategy typically has the goal of inducing broad HI titers elicited against the HA head from strains within a subtype (65)(34)). Vaccines against M2 are also being considered since this protein is relatively conserved. Vaccines targeting NA have also been considered since NA antibodies can be broad within a particular subtype. The major hurdle will be to not only elicit immune responses that are broadly protective, but also durable and long-lasting, otherwise a universal flu vaccine will not be of much use.

For Chapter 2, work will be presented on developing a mosaic nanoparticle that co-displays up to 8 different trimeric HAs from a set of Group 1 and Group 2 strains. These mosaic nanoparticles were compared to a corresponding admixture of homotypic particles. The breadth of antibody responses elicited by these particles was characterized along with cross-reactive B-cell responses.

## **Part B: Coronaviruses**

### **The SARS Outbreak:**

In late 2002 reports of a mysterious respiratory illness emerged from China with unknown origin. By mid March 2003 new outbreaks were occurring in other parts of Asia, and although the cause of these outbreaks was still unknown, the WHO put out an alert, with information on disease severity, transmission, and death rate all unclear (76, 77). The first transmission events happened in Foshan City, Guangdong Province, China, on November 16, 2003. From November 2002 to February 2003 there were a total of 305 cases (76, 77). An index case arrived in Hong Kong by February 22, 2003 when an infected medical doctor checked into the 9th floor of the Metropolitan hotel, with a total of 10 infections being linked to that case. Eight of the 10 guests infected resided on the 9th floor. The index patient apparently vomited, and the vomit was subsequently vacuumed, and therefore it is possible that resulted in the production of aerosols. That single floor is the setting responsible for seeding the global spread that ensued, with the 10 hotel guests and visitors carrying the virus to hospital systems in Hong Kong, Toronto, Singapore, and Vietnam, triggering deadly outbreaks (76, 77).

The Amoy Gardens apartment complex in Kowloon Bay, Hong Kong was the setting of another interesting but frightening outbreak (76, 77). A 33-year-old man with renal disease was visiting his brother, who lived in a 15-story high rise block in that complex, after getting treated at a hospital nearby on March 14, 2003. When visiting his brother he had diarrhea and used the toilet (76, 77). Within the following month 321 cases were linked to Amoy Gardens, with 41% of the cases linked to the block the brother lived in, and what is now known as the point source outbreak (76, 77). The secondary cases were suggested to be caused by contaminated sewage water which became aerosolized when toilets were flushed and bathroom exhaust fans were turned on (76, 77).

In the hospital setting health care workers and patients were vulnerable as proper safety precautions were not employed since the transmission mechanisms were unknown. This resulted in rapid spread of this virus in hospitals, known as nosocomial transmission, with numerous health care workers, patients, and visitors falling ill with often deadly consequences (76, 77). These types of cases were known as superspreader events. Secondary cases in homes by contrast besides the

Amoy Gardens or Metropolitan hotel outbreak were low with rates of around 15 % in Hong Kong and 6% in Singapore (76, 77)). Interestingly, secondary cases from airlines were also minimal with only 29 cases total linked to air travel, with some exceptions. For example, one symptomatic case flying from Hong Kong to Beijing resulted in 22 out of the 119 passengers getting infected (76, 77)).

A total of 8096 cases were reported by July 2003, with 774 deaths from outbreaks in 27 countries (78). The outbreak was declared over at this point though four additional cases occurred December 2003-January 2004 most likely from zoonotic transmission, with no new cases declared after 2004 (78).

The attack rate of children is lower than for adults. Children also had less severe disease, with adolescents having similar illness to adults but lower case-fatality ratios (CFRs). Symptoms in children included fever, cough, rhinorrhea, myalgia, chills, and headache (76, 77)). SARS cases in children did feature abnormal chest scans (computer assisted CT or radiograph) with infiltrates and opacities suggesting inflammation in the lungs.

The incubation period for SARS-CoV is 2-10 days, with 4-6 days being the median (76, 77)). In adults in the first week patients present with fever, malaise and headache (76, 77). By week 2 cough, dyspnea (trouble breathing) and diarrhea appear (76, 77). Towards the end of the 2nd week, some patients deteriorate, with 20% requiring intensive care (76, 77). This deterioration is accompanied by laboratory confirmed conditions such as lymphopenia (loss of lymphocytes), thrombocytopenia (low platelet count). Chest radiographs or CT scans will show abnormalities such as ground glass opacification (patchy white parts suggesting fluid and cells in the lungs). In fatal cases several features are prominent, epithelial cell proliferation (resulting in scarring of lung tissue), diffuse alveolar damage (damage of lung tissue), macrophage infiltration of the lungs, and pleural effusion (fluid buildup in lungs) (76, 77). Comorbidities such as old age, cardiac disease, and lung disease correlate with fatal outcomes and disease severity.

The virus responsible was identified by March 2003 as a novel coronavirus known as SARS-CoV. The virus was able to be sequenced after it was found to replicate in VERO cells (African

Green Monkey cell line), with multiple labs subsequently obtaining the sequences of the viral genome, which was at the time difficult due to the large size of the genome (~30kb) (76, 77). These sequences resulted in the development of diagnostics to track infection, as well as tease the origin (76, 77)).

Almost identical SARS-CoV-like viruses (99.8% nucleotide identity to SARS-CoV) were isolated from Himalayan palm civets (*Paguma larvata*) and raccoon dogs (*Nyctereutes procyonoides*) in a live animal market from Shenzhen, China (76, 77)). ORF8 sequences from infected palm civets matched the ORF8 sequences from the early cases. As the outbreak progressed a 29 nucleotide deletion in ORF8 emerged which became the predominant variant in the middle and late parts of the outbreak (76, 77)). Therefore, palm civets were presumed to be the source of the outbreak. Furthermore in the animal market 40% of wild animal traders and 20% of those who slaughter the animals and 5% of vegetable traders in the same market were seropositive for the SARS-CoV, further suggesting that palm civets in wild animal markets are the source of the outbreak (76, 77)).

### **Discovery of SARS-like Coronaviruses**

A couple of years after the 2002-2003 SARS outbreak bats were identified as the natural reservoir for SARS-like coronavirus (79). SARS-CoV and palm civet SARS-like CoV nestled in with a swath of these SARS-like coronaviruses that were isolated from different bat species in southern China (79). Following this SARS-like coronaviruses were isolated from bats in Asia, Europe and Africa (80). However none of these bats were directly related to SARS-CoV and therefore were not believed to be the progenitors of the SARS-CoV outbreak (80). In 2013 whole genome sequences of two SARS-like viruses were isolated from horseshoe bats that were found in a cave in Yunnan Province, China. These viruses, labeled RsSHC014 and WIV1-CoV, were the most closely related viruses to SARS-CoV identified (80). WIV1-CoV in particular was isolated through the passage of bat fecal samples in VERO cells where cytopathic effects were seen (80). The isolated WIV1-CoV

was able to use horseshoe bat, civet, and human ACE2 for cellular entry based on immunofluorescence and RT-PCR assays (80). This started to suggest that other SARS-like coronaviruses with human outbreak potential existed (80).

To further characterize the infection potential of these SARS-like coronaviruses chimeric virus that contained the SHC014 spike protein on a mouse adapted SARS-CoV backbone was used to tease out the infection potential of SHC014-CoV (81). The chimeric SHC014 virus was capable of infecting and replicating in both Calu-3 cells (lung cell line) as well as human airway epithelial cells (HAE), suggesting that it has human tropism (81). Furthermore, infection of mice with the chimeric SHC014 did result in infection which resulted in some disease characterized by modest weight loss, and minimal lethality in aged mice, and less antigen staining in the lung. SARS-CoV infection on the other hand did result in severe weight loss lethality and extensive antigen staining (81). The chimeric SHC014 was not able to be neutralized by SARS-CoV neutralizing monoclonals, suggesting limited neutralizing cross reactivity (81). Finally, the full length SHC014-CoV was cloned through reverse genetics. The authentic SHC014-CoV was able to infect VERO cells, but did not replicate as well in HAEs with respect to SARS-CoV. Finally SHC014 was able to infect mice, though not as well as SARS-CoV (81). This suggests that SHC014 is viable for cross-species infection though it may need more adaptation to become more pathogenic and transmissible (81).

A related study was carried out using WIV1, which shares homology with SARS-CoV and has 11 out of the 13 ACE2 contact residues conserved (82). WIV1 full length authentic virus was produced with a reverse genetics approach (WIV1-CoV). A WIV1 chimeric virus (WIV1 spike with SARS-CoV backbone) was also produced that contained mouse adapting mutations (WIV1 MA15) (82). The WIV1 MA15 and WIV1-CoV were both able to infect and replicate in VERO and HAEs to the same degree. WIV1 MA15 virus as well as mouse adapted SARS-CoV MA15 were used to infect wt mice (82). Infection of the chimeric WIV1 was attenuated in mice with virtually no weight loss and lower viral titers in the lungs than SARS MA15. Authentic WIV1 infection in hACE2 mice

was also carried out, and WIV1 infection was attenuated with lower weight loss and mortality accompanied by lower viral titers in the lung than SARS-CoV. SARS-CoV neutralizing antibodies were however cross-neutralized WIV1 (82). The similar attenuation feature of WIV1 infection that was seen in SHC014 suggests that WIV1 is viable for cross-species transmission but it will require more adaptation for disease (82).

In an attempt to see if any of these SARS-like coronaviruses have infected humans, serology studies were carried out in 218 residents of four villages in Yunnan Province, China, located near two caves where SARS-like coronaviruses have been isolated from horseshoe bats. Occupational most of the villagers were farmers with constant contact with livestock. Notably, 9% of the participants witnessed bats flying around nearby, and one had even handled a bat corpse. 3% (6/218) of the serum samples collected were positive for binding to the NP of the SARS-like CoV Rp3 via ELISA as well as western blot. This serum reactivity was confirmed by western blot analysis. The serum did not neutralize SARS-CoV and WIV1-Cov, and it did not bind to the RBDS of SARS-CoV and WIV1-Cov either, suggesting another SARS-like CoV is responsible for the serum reactivity (83). All of the seropositive residents have seen bats flying around. Furthermore, two of the seropositive residents had travel history to major cities like Kunming and Shenzhen. Finally, none of the seropositive residents had reported any recent illness. Altogether this study suggests that spill-over is taking place, though it is an infrequent event.

### **The COVID-19 Pandemic:**

On December 31, 2019 the WHO was alerted to a cluster of pneumonia cases of unknown origin, in Wuhan, China (84). Less than two weeks later the entire genome of the virus was deposited to GenBank, with the cause being a new coronavirus, labeled severe acute respiratory syndrome coronavirus-2 or SARS-CoV-2, related to SARS-CoV and the SARS-like Coronaviruses (84). The disease that it caused was labeled coronavirus disease 2019 or COVID-19.

The first actual SARS-CoV-2 cases were reported in December 2019 in Wuhan, China, where from December 18, 2019 to December 29, 2019 five patients were hospitalized and one

patient died. By December 29th, the source of many of the cases was linked to the Huanan Seafood Wholesale Market (85). However some of the earliest cases where the onset of illness occurred as early as December 7th were not associated with the market (85). By January 1, 2020, 47 patients were admitted and identified as having clinically-confirmed pneumonia, with emergency responses being initiated by the Chinese Center for Disease Control (CCDC) and the sequencing of the samples taking place (85). By December 8, 2020 a new coronavirus was confirmed to be the causative agent, and by December 11, 2020 the virus sequence was deposited into GenBank. By December 15th, 2020 strict screening measures were employed for individuals exiting Wuhan where infections were in the thousands accompanied by many reported deaths. By December 19th, 2020 the first case was reported outside of Wuhan in another region of China (85). On January 23, 2020 the entire city of Wuhan and its 9 million residents were put under a strict lockdown, with most of the Hubei Province being on lockdown the following day, and other travel restrictions employed across China (86). Some of the first cases outside of China were identified mid-January in Thailand and Japan, both being linked to travel from Wuhan. In the ensuing weeks cases were reported exponentially in many countries, first directly related to travel to Wuhan, but with many unrelated. Large and deadly outbreaks followed in many regions and countries, and the WHO declared the virus as a pandemic on March 3, 2020 (87). The pandemic spread has emerged as waves of infections with three peaks. As of May 2021, over 150 million people have been infected, with 3.2 million deaths (88).

One of the things that sets the SARS-CoV-2 virus apart from previous coronavirus outbreaks in the 21st century is the efficient person to person transmission, with an  $R_0$  estimated to be as high as 5.7, an incubation period of average 5.7 days, with peak viral loads corresponding to right before symptom onset (89, 90). This accompanies many reports of asymptomatic infections or presymptomatic transmission (48% to 62%) being one of the major drivers of SARS-CoV-2 spread (91). It was initially considered that SARS-CoV-2 spread through respiratory droplets (particles 5  $\mu$  m to 10

$\mu\text{ m}$ ), however aerosol spread ( particles  $< 5 \mu\text{ m}$ ) was later believed to be a large contributor to the outbreak with 48% to 62% of transmission believed to be from aerosols and many human activities having been shown to produce aerosols as well (91). Fomite transmission ( contact with contaminated surfaces) is believed to be minimal (91).

Shutdown measures across China were largely successful at blocking transmission. Many other nations around the world employed less stringent lockdown measures or non-pharmaceutical interventions (NPI), such as closing international travel, closing schools, cancelling public events, stay at home requirements, closing workplaces and restricting gatherings (92). These NPIs had varying success rates at reducing transmission. In addition, masks or face coverings have shown some efficacy in reducing transmission as well, and were mandated in many different countries and regions (91).

Covid-19 disease has a wide severity range with 81% of individuals having mild manifestations, 14% with severe manifestations, and 5% with critical manifestations (respiratory failure, shock, and multiple organ failure) (93). Within the first week of symptom onset, known as the viral stage, symptoms could include fever, dry cough, shortness of breath, nausea, vomiting, diarrhea, myalgia (pain), headache, fatigue, and rhinorrhea. Loss of sense of smell and taste (anosmia and ageusia) is very common and in some cases is the only symptom as well. For mild infections symptoms typically resolve (93). There are also reported asymptomatic cases, though the percentage of asymptomatic cases does vary, and it is unclear whether they are truly asymptomatic (93). Along the second week, conditions can deteriorate rapidly in severe and critical patients, with many complications arising. Such complications include hypoxemia (low  $\text{O}_2$  levels in the blood) and respiratory failure which can require mechanical ventilation, impaired function of the heart, brain, lung, liver, kidney, and arterial thromboembolic events (blood clots) (93). A common abnormality for COVID-19 is the presence of diffuse, ground-glass opacities in chest scans of the lung (suggesting fluid buildup in the lung), and it was described in the earliest patients at the onset of the outbreak. Due to the immunopathological origins of the severity of the 2nd week of disease, typically steroids are

prescribed such as dexamethasone, which have shown some efficacy. Comorbidities such as heart disease, diabetes, and kidney disease among others have all been associated with increased risk in disease severity (93). In the US the case fatality rate varies dramatically by age with 0.3 deaths/1000 cases for young children and adolescents (5-17) to 305 deaths/1000 cases for those aged 85 and older, with the >65yo age group being considered the highest risk (93).

The emergence of new SARS-CoV-2 variants has been a cause for concern (known as variants of concern or VOC), especially with efforts to control the pandemic spread through NPIs and vaccines (94). These variants include B.1.1.7 (first discovered in the UK), B.1.351 (first identified in South Africa), B.1.427 (First identified in California), and B.1.526 (identified in California), P1 (identified in Brazil), and B.1.167 (recently identified in India). There is evidence to suggest that some of these variants emerged as escape mutants from antibody selection pressures (94). For instance the B.1.351, B.1.526, and P1 all have the E484K mutation in the S protein whereas B.1.167 has a E484Q mutation in the S protein, both of which have been shown to abolish binding for a class of SARS-CoV-2 antibodies (95). There is also evidence that some of these variants resulted in increased transmission, such as in the case for B.1.1.7, where it became the dominant strain with new infections in the UK late 2020, though this may have also been due to non-adherence to social distancing measures (96).

#### **Coronavirus structure and genome:**

Coronaviruses (CoVs) are a large and diverse class of enveloped positive-sense single stranded RNA viruses belonging to the order *nidovirales*, the suborder *coronavirinae*, and the family *coronaviridae* (97). The *coronaviridae* family can further be classified into the *ortocoronavirinae* order, which is divided into four genera: *alphacoronaviruses*, *betacoronaviruses*, *gammacoronaviruses*, and *deltacoronaviruses* (97). The betacoronavirus family, which will be the main focus of this

intro, can be further divided into lineages: Sarbecoviridae (SARS-like betacoronavirus), hibecoviridae, merbecoviridae (MERS-like betacoronavirus), nobecoviridae, and Enbecoviridae (HKU1 and OC43). The sarbecovirus family includes the severe acute respiratory syndrome coronavirus (SARS-CoV) and severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) and the merbecovirus family contains Middle Eastern respiratory syndrome coronavirus (MERS-CoV).

Alphacoronaviruses and betacoronaviruses exclusively infect mammalian species, whereas gammacoronaviruses and deltacoronaviruses infect a wide range of vertebrates including birds and reptiles (97). In humans HCoV-229E (alphacoronavirus) and HCoV-OC43 (betacoronavirus) have been circulating together for at least a century along with the more recently discovered HCoV-NL63 (alphacoronavirus) and HCoV-HKU1 (betacoronavirus). These viruses cause mild infection of the upper respiratory tract in humans with “common cold” like symptoms (97).

The three betacoronaviruses SARS-CoV, SARS-CoV-2, and MERS-CoV in contrast are far more pathogenic in humans and have emerged over the last two decades. They can infect bronchial epithelial cells and the pneumocytes of the lower respiratory tract as well as the upper respiratory tract which can result in severe and fatal respiratory pathology as well as acute lung injury (ALI), with little therapeutic interventions available (97).

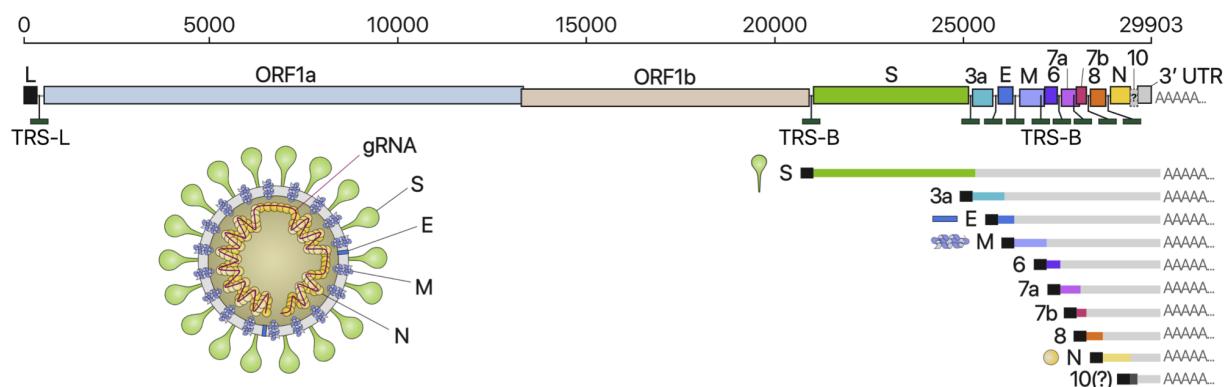


Figure 6: SARS-CoV-2 Genome. Taken from (98).

The coronavirus genome is large (> 30 kb), and contains long non-translated regions (NTRs) flanking the 5' and 3' end that form secondary RNA structures necessary for RNA replication (97). The coronavirus genome is also capped and polyadenylated. The 5' end contains the ORF1a and ORF1b reading frames, which cover two-thirds of the viral genome. The ORF1a and ORF1b encode for 15-16 non-structural proteins (nsp). 15 of the nsps make up the replication and transcription complex, which include functions such as the RNA polymerase, RNA processing, RNA modifying, and RNA proofreading function, among others, which help to maintain the integrity of the large RNA genome. The remaining ORFs encode structural and accessory proteins located at the last 3' end (last third of the genome). These ORFs are also transcribed as a nested set of subgenomic mRNAs (sgRNA). A map of the viral genome is shown in figure 1.

The coronavirus structure contains Spike (S), Envelope (E), Membrane (M), and for some betacoronaviruses hemagglutinin esterases (HE) all embedded in the host-derived membrane of which makes the viral envelope. Inside the virus envelope there is the nucleocapsid protein (N) which encapsulated the (+) single stranded RNA genome, with M and E ensuring the incorporation of the N encapsulated RNA genome in the particles. The S protein sticks out of the envelope and is the protein that attaches to host receptor and mediates cellular entry and fusion. The coronavirus accessory proteins are variable between viruses, with limited conservation, and because of this their function is largely unknown for most viruses. The accessory proteins are thought to modulate host immune responses to infection and can be determinants of viral pathogenesis.

### **Coronavirus life cycle**

The first step of viral entry involves attachment to host cells and cell entry and is mediated by the interaction of coronavirus spike protein (S) with cell entry receptors. The cell receptors recognized by S have been identified as human angiotensin converting enzyme-2 or ACE2 (for SARS,

SARS-2, and NL63), dipeptidyl peptidase 4 or DPP4 (for MERS), and aminopeptidase N or APN (for NL63). The distribution of receptor expression in different tissues guides the tropism as well as pathogenesis for coronaviruses (97). Intracellularly upon release of the genome into the cytoplasm, the full genome is replicated and packaged into new virions.

The release of the coronavirus genome into the cytoplasm starts a complex process of viral gene expression that is spatially and temporally regulated. ORF1a and ORF1b are directly translated first from the genomic RNA to produce ppa1a and ppa1ab (made from a programmed ribosomal frameshift at the ORF1a ORF1b overlap) (97). The stoichiometry of the pp1a and pp1ab efficiency of the frameshift between ORF1a and ORF1b, which was determined by ribosomal profiling to be 45-70% for SARS-CoV-2. Therefore, ppa1a is 1.4-2.4 times more expressed than pp1ab. A total of 16 non-structural proteins are released upon pp1a (nsp1-11) and pp1ab (Nsp1-10 and Nsp12-16) proteolytic cleavage by nsp3 (papain-like protease; PLpro) and nsp5 (chymotrypsin-like protease) (97). Nsp5 is also known as the 3C- like protease (3CLpro) due to its similarity to the picoviral 3C proteases, as well as main protease (Mpro) because it is the protease that predominantly cleaves most of the nsp cleavage sites. Nsp1 apparently is released first where it immediately targets host-cell translation by promoting host-cell mRNA degradation as well as targeting the 40S ribosomal subunit (99, 100). Interestingly when nsp1 binds the 40S ribosomal subunit it prevents the binding of cellular mRNAs by blocking the mRNA binding groove, hence shutting down cellular translation. Viral mRNA with the 5'UTR is better and being translated than cellular RNA, so it is believed that the nsp1 levels are attenuated such that viral mRNA will outcompete host cell mRNA for translation (99).

Nsp2-16 compose the viral replication and transcription complex (RTC), and are localized to defined subcellular compartments where they interact with host-cell factors and orchestrate the viral replication cycle (97). Nsp2-11 play the necessary supporting role for the RTC such as host-immune evasion, providing replication cofactors and intracellular membrane modulation. Nsp12-16 contains the core enzymatic function for the RTC, including RNA polymerase, RNA modification,

and importantly RNA proofreading. Nsp12 is the RNA-dependent RNA polymerase (RdRp), with nsp7 and nsp8 (primase or a 3'-adenylyltransferase) as its cofactors (97). Nsp14 is the unique RNA proofreader for coronaviruses and has 3'-5' exonuclease activity. Coronaviruses also have capping machinery, with nsp13 as the 5' triphosphatase activity, ns14 and the N7- methyltransferase, nsp16 as the 2'- O- methyltransferase, and nsp10 as a cofactor with others not fully elucidated including the guanylyltransferase, a key enzyme involved in the formation of the 5' cap (97)).

The nsp5 or Mpro protease is a critical protein in releasing the majority of the nsps from the polyprotein which is crucial for the viral life cycle. Its cleavage function is sequence specific and can easily be mimicked with a peptide that resembles the cleave site. Structures have been determined of the Mpro in complex with inhibitory peptides that were computationally designed and allowed for screening of more compounds resulting in novel leads for Mpro based inhibitors (101). These could serve as a new set of therapeutics against SARS-CoV-2 with some already showing strong antiviral activity in animal models (102).

Viral replication begins with the transcription of the genomic RNA into a negative stranded template, which gets copied into new genomic RNA strands. The newly synthesized gRNAs get either packaged into new virions or used for translation of more nsps and RTCs. The hallmark of the entire nidoviraleae family of viruses is the production nested set of 3' and 5' co-terminal subgenomic RNAs (sgRNAs) produced through discontinuous transcription (97)). The viral genome is interspersed with transcription regulatory sequences (TRS) located upstream of most ORFs, especially the ones in the 3' third of the genome. Transcription of the negative stranded RNA copies are interrupted by these TRS elements with the 5' leader TRS at the 5' end of the genome interacting with the 3' end of the TRS on the negative strand being transcribed (97)). The end result is the production of negative stranded sgRNAs that serve as templates to produce positive stranded sgRNAs which then get translated into accessory and structural proteins. Although the sgRNAs are structurally polycistronic, by practice they behave as monocistronic RNAs, where only the 5' ORF gets translated (97)). There is also the presence of non-canonical RNA products from discontinuous

transcription. These RNA products may also be derived from recombination events (98). This may be a mechanism leading to the diversification of coronaviruses and production of variants, possibly creating novel ORFs or accessory proteins that would enhance viral replication or better dampen host immune responses (97)).

The nsp12 RdRp accounts for the main enzymatic activity of the RTC and is therefore an attractive drug target. The structure of the SARS-CoV-2 nsp12 and the cofactors nsp7 and nsp8 has been elucidated and shown to have remarkable conservation to SARS-CoV (aa identity is >95%) (97). Due to the high similarity of the RdRp of SARS-CoV-2 with other coronavirus RdRps as well as other viral RdRps, drugs such as remdesivir (RDV) have been repurposed for use in treating SARS-CoV-2 infections. RDV competes for ATP in the active site of the SARS-2-CoV RdRp, which results in delayed chain termination (97).

The coronavirus structural proteins S, E, M, and N assist in budding of new virions into the ER to Golgi compartment ERGIC, and the virions are then thought to be exocytosed to the plasma membrane (97). There is evidence however that viral egress occurs via the lysosomal trafficking pathway as well (97).

To date ORFs encoding five accessory proteins have been identified: ORF3a, ORF6, ORF7a, ORF7b, and ORF8, with potentially ORF3b and ORF9b as well as ORF10 being other possible accessory proteins (97). Although between different coronavirus lineages ORF accessory proteins are not conserved and are highly variable, within a viral species they are relatively conserved and play an important role in replicating in the natural host. There are some differences between SARS-CoV and SARS-CoV-2 ORFs. For instance ORF3b in SARS-CoV-2 is much shorter than in SARS-CoV, with possible implications for it as an interferon agonist in its truncated form (97)). ORF8 between SARS-CoV and SARS-CoV-2 shows low sequence homology, where in SARS-CoV the ORF8 underwent a 29 nucleotide deletion in the middle and late stages of the SARS outbreak, which broke apart ORF8 into two truncated gene products ORF8a and ORF8b and re-constitution of ORF8 leads to a slightly increased cell culture fitness (97). SARS-CoV-2 which has

a full length ORF8 may lead to a downregulation of MHC1 and therefore be involved in immune evasion. However, a deletion of ORF8 was seen in an isolated set of cases in Singapore. Altogether removal of ORF8 may play a role in viral attenuation in humans (97).

Another hallmark of the coronavirus replication cycle is the formation of replication compartments or replication organelles, which are thought to be formed by both viral and host factors (97). Specifically nsp3, nsp4, and nsp6, which are membrane bound, have been implicated in diverting host endomembranes into these replication compartments (97). Coronavirus infection results in the formation of ER-derived perinuclear double membrane structures such as double membrane vesicles (DMVs) and double membrane spherules, and these organelles are highly dynamic. Parts of the RTC such as nsp2, nsp5, and nsp6 are shown to be tethered to the DMVs, it is still unclear exactly where replication takes place (97). Interestingly the dsRNA intermediates formed from viral replication are exclusively located in the DMVs, suggesting a role in immune evasion, and preventing the dsRNA intermediates from being recognized by cytosolic RNA sensors. Viral RNA synthesis for coronaviruses has been shown to take place only in the DMVs, and recently pores have been shown to exist that may help in release of the viral RNA for packaging and translation (97).

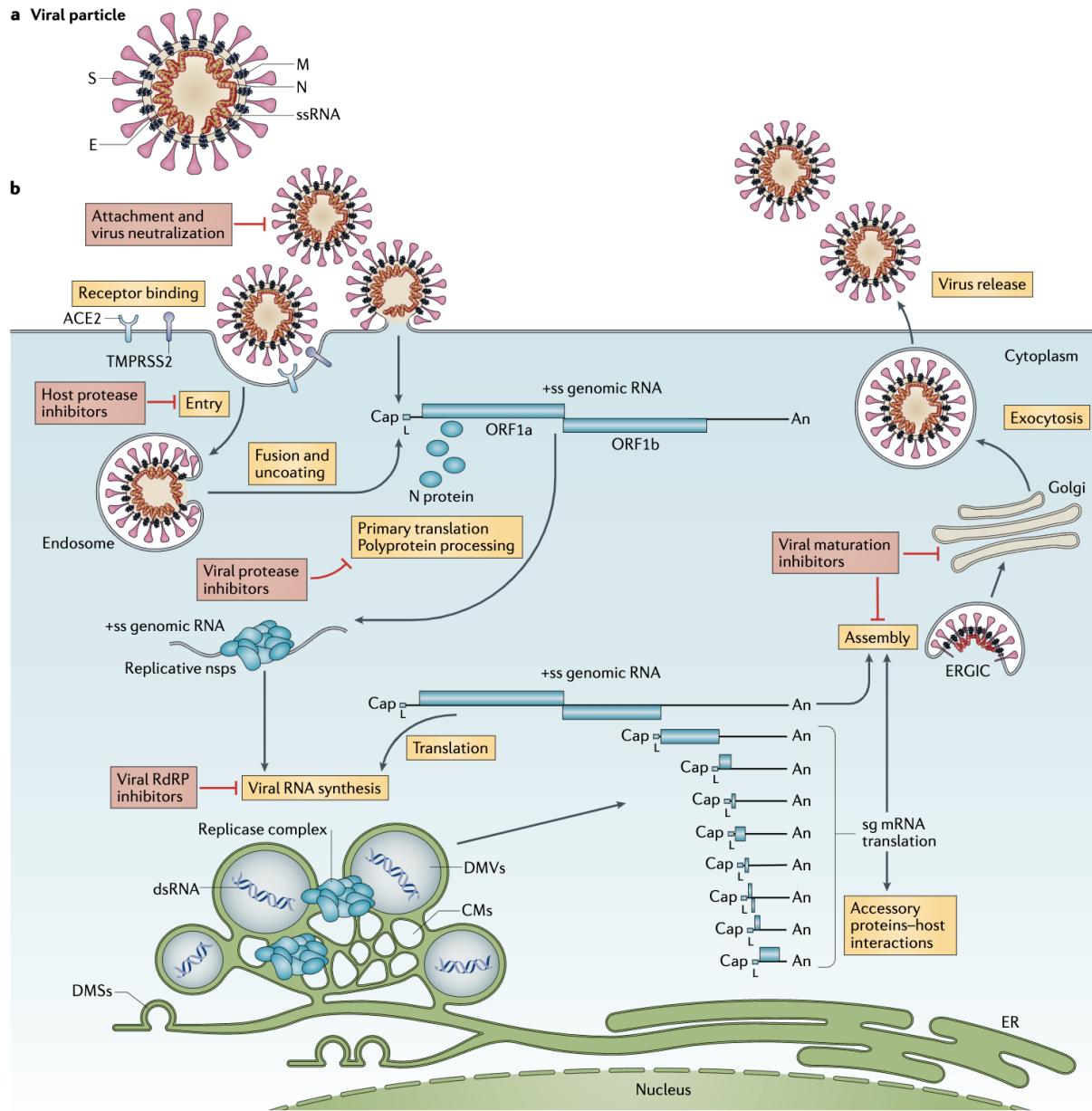


Figure 7: Coronavirus replication cycle. Taken from (97).

**The Coronavirus Spike Protein and Its Effect on Viral Tropism, Transmission, and Pathogenesis**

The coronavirus spike is a class 1 fusion homotrimeric protein that is divided into two structurally and functionally distinct subunits: the S1 and the S2 (97). The S1 subunit contains the receptor binding domain, which functionally is what binds to the host cell receptor, and therefore the RBD is the determinant of viral tropism. The RBD of SARS-CoV-2 and SARS-CoV can exist in two distinct conformations, designated as the “up” and “down” conformation (103). In the up or open conformation, the RBD is pointed up with the receptor binding motif exposed for ACE2 binding, whereas in the closed or down confirmation the RBD is facing down on the trimer apex, with the receptor binding motif occluded. The S2 domain contains two heptad repeats as well as the fusion peptide which mediates viral membrane-host membrane fusion. Binding of the RBD to the host receptor mediates conformational changes which rearrange the S2 domain so that it can mediate viral membrane fusion.

The ACE2 receptor was identified as the receptor for SARS-CoV entry shortly after the SARS outbreak determined by syncytia formation between spike and ACE2 expressing cells (104). Following the beginning of the COVID-19 pandemic ACE2 was determined to be the receptor for SARS-CoV-2 entry as well determined by showing the dependence of a SARS-CoV-2 S pseudotyped VSV on ACE2 for cell entry (104, 105). The use of ACE2 as the RBD was verified with high resolution structures of the S and as well as the RBD complexed with ACE2 (103, 106, 107). This was also predicted to be the case due to the sequence homology of the SARS and SARS-2 spike (78% sequence homology) which is particularly high though with distinct changes when it comes to the ACE2 contact residues in the receptor binding motif of the RBD (104). There are other species of bat sarbecoviruses that are also ACE2 tropic, which also maintain the RBD-ACE2 contact residues (80, 81, 105). The closest relative of SARS-CoV-2 is the bat virus RaTG13 (96% aa identity for spike, and 96.2 nucleotide identity for the entire genome) but it has different residues at the SARS-CoV-2 ACE2 contact positions (with only one identical residue out of the six critical ACE2 contact residues). It should be noted that RaTG13 S is able to mediate entry via human ACE2. Furthermore ACE2 tropic SARS-like viruses have broad tropism for ACE2 from many different

mammalian species, most likely due the conservation of RBD binding site on ACE2 (108, 109). Altogether it is suggested that the ACE2 binding trait of SARS-CoV-2 was not one recently acquired by recombination but rather an ancestral trait that has been shared with similar viruses (87).

The SARS-CoV-2 virus appears to have arisen via frequent recombination events with related circulating sarbecoviruses in bats. Based on phylogenetic dating methods that took into account recombination events, divergence between SARS-CoV-2 from the sarbecovirus reservoir has been estimated to have occurred between 2-7 decades prior to its arrival in the human population in 2019, suggesting that SARS-CoV-2 has been circulating in bats unnoticed for decades (87). The SARS-CoV-2 Spike can be divided into three parts: the N-terminal domain (NTD), the middle receptor binding domain (RBD), and the S2 (87). The RBD middle portion of the S from SARS-CoV-2 is closely related to SARS-CoV as well as the ACE2 tropic strains, WIV1-CoV and RsSHC014-CoV. In contrast the NTD and the C-term S2 portion of the SARS-CoV-2 spike comfrom ZXC21-CoV and ZC45-CoV (87).

# Spike RBD

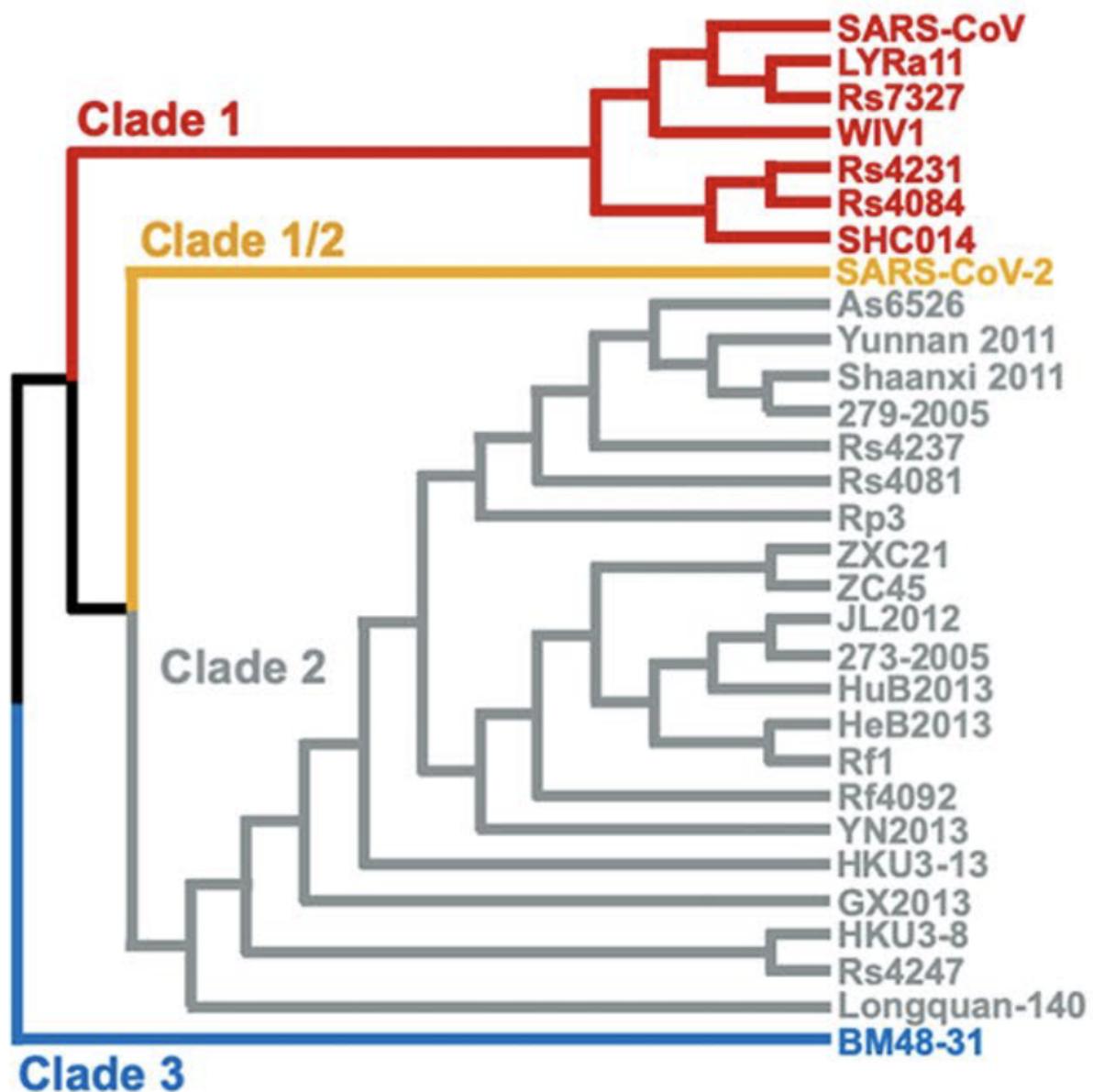


Figure 8: Sarbecovirus RBD phylogenetic tree. Taken from (105)

Based on the RBD sequence identity the sarbecovirus lineage, aka lineage B or the SARS-like coronaviruses, appears to be divided into four clades: clade 1, which includes SARS-CoV, WIV1-CoV, and RsSHC014-CoV; clade 1/2 which includes SARS-CoV-2; clade 2 which includes ZC45 and ZXC21; and clade 3 which includes BM-4831 (105). In an effort to identify receptor usage by different sarbecoviruses, VSV pseudoviruses were pseudotyped with chimeric spikes, containing the RBDs from 29 different sarbecovirus strains on a SARS-CoV spike backbone, and analyzed for entry into different human and mammalian cell types (105). Most of the clade-1 strains were permissive to VERO cells, human tropic, and dependent on ACE2 for viral entry (with SARS-CoV-2 of clade 1/2 also being ACE2 tropic) (105). The clade 1 mediated entry was enhanced by prior protease treatment. Interestingly some clade 2 strains were capable of viral entry into VERO cells as well as GI tract (Caco-2) and liver (Huh-7) cell lines and this entry was dependent on prior protease treatment. Clade 2 and clade 3 strains were not dependent on ACE2 for viral entry, suggesting that some other receptor on the VERO and human cell lines were mediating viral entry (105). Finally, chimeric clade 2 and clade 3 spikes that contained a consensus receptor binding motif from clade 1 strains gained the ability to recognize ACE2 for viral entry of pseudoviruses. Altogether, there is an ability of sarbecoviruses to use receptors on a wide variety of mammalian cells for cellular entry, and protease compatibility most likely drives the tropism of these viruses (105).

Similar to influenza, another important feature that dictates viral tropism is the proteolytic cleavage of the spike by host derived proteases. For SARS-CoV spike, TMPRSS2 has been identified as the main host proteases required for spike cleavage, though the endosomal cysteine protease cathepsins (CatB and CatL) appear to be utilized as well (110, 111). The two cleavage sites were determined to be the S1-S2 (R667) cleavage site, located between the S1 and S2 subunits, and the S2'(R797) cleavage site located right before the FP. Mutating both the S1-S2 and S2' cleavage sites inhibited the efficient cleavage of the spike as well as trypsin mediated pseudoviral entry and syncytia formation (112). TMPRSS2 is expressed along the human respiratory tract and ap-

pears to be important for the viral spread and pathogenesis of SARS-CoV (113). TMPRSS2 deficient mice models appear to have decreased body weight loss and a lower viral load when infected with SARS-CoV (113).

TMPRSS2 was also determined to be the predominant host protease utilized by SARS-CoV-2, with CatB/L playing a more limited role (114). Blocking of TMPRSS2 with the inhibitor camostat was able to block viral entry of Calu-3 lung cells as well as human primary airway epithelial cells (114). Viral entry of SARS-CoV-2 was not blocked in VERO cells (African green monkey cell line) by the TMPRSS2 inhibitor whereas E64D (a cathepsin inhibitor) did block viral entry (114). However, VERO cells expressing TMPRSS2 were inhibited more by TMPRSS2 inhibitors than by cathepsin inhibitors. Altogether this suggests TMPRSS2 is the preferred host protease for SARS-CoV-2 viral entry (114). This preferred host protease usage appears to be the reason why hydroxychloroquine was not proven to be effective in SARS-CoV-2 treatment (115). Hydroxychloroquine works in part by preventing the acidification of the endosome which is required for the activation of the cathepsin B/L, therefore preventing these proteins from being utilized for viral entry (115). Hydroxychloroquine worked in inhibiting viral entry only when TMPRSS2 was inhibited or suppressed, where complete inhibition required both proteases to be blocked. Therefore SARS-CoV-2 can utilize either TMPRSS2 or CatB/L for viral entry, which suggests that SARS-CoV-2 can enter cells either directly on the plasma membrane or through the endosomal pathway, with direct entry possibly being the preferred way. This also speaks to the use of TMPRSS2 inhibitors in a clinically relevant manner (116).

Based on the studies highlighted above, it appears that recombinatorial events guide the diversification of coronaviruses and drive the zoonotic emergence of coronaviruses into new species. This highlights the importance of viral surveillance that can further elucidate the viral diversity for sarbecoviruses in the wild and determine which of these viral strains are capable of human tropism. Increased surveillance is therefore a necessary part of pandemic preparedness and what

needs to be done in order to prevent or at least minimize the severity of another COVID-19-like outbreak or pandemic.

Although both SARS-CoV and SARS-CoV-2 use ACE2 in entry, they display marked differences in replication efficiency, spread, and transmission dynamics (97, 116). SARS-CoV mainly infects various respiratory epithelial cells in the lung, predominantly type II pneumocytes as well as alveolar macrophages, which matches the highest levels of surface ACE2 expression, which is in alveolar epithelial cells in the respiratory tract (117, 118). For SARS-CoV-2, infection seems to be in both the upper respiratory tract as well as the lower respiratory tract, suggesting that the SARS-CoV-2 virus is capable of broader tissue tropism (119). It is the ability to replicate efficiently in the upper respiratory tract that may allow for more efficient transmission of SARS-CoV-2 with respect to SARS-CoV (91).

One of the reasons for the difference in host-cell tropism for SARS-CoV and SARS-CoV-2 may be due to the receptor affinity of the S protein to the ACE2, with greater affinity for the ACE2 receptor correlating with disease severity of SARS-CoV infections (97). The SARS-CoV-2 RBD on its own recognizes ACE2 with approximately the same affinity as SARS-CoV. However the SARS-CoV spike actually recognizes the ACE2 with a slightly lower affinity than SARS-CoV, which might be explained by the RBD on SARS-CoV-2 being more in the closed confirmation (103). This may be supported by the fact that the SARS-CoV-2 D614G variant that emerged and dominated relatively earlier in the pandemic featured higher pseudoviral entry as opposed to the wt Wu-1 original variant (120). The affinity of the D614G variant S protein to ACE2 was slightly higher than that of the wt, suggesting that the RBD might be more in the open conformation (120). In addition for SARS-CoV-2 to RBD accessibility, Neuropilin-1 (NRP1) has been determined to be a host factor for viral entry (121). The S1-S2 cleavage of the SARS-CoV-2 spike by furin proteases exposes a RRAR CendR motif which has been shown to bind to NRP1 (121). Furthermore, blocking NRP1 S1 interaction results in a loss of infectivity in cell culture. It should be noted that NRP1 is expressed on most respiratory epithelial cells (121).

One of the most notable differences between SARS-CoV-2 and the other sarbecoviruses is the presence of the polybasic cleavage site (PRRARS) in the S1-S2 boundary. This multibasic cleavage site allows for the efficient cleavage of the SARS-CoV S protein by ubiquitously expressed furin proteases (103). The addition of the furin cleavage site in SARS-CoV-2 was most likely a critical event in its evolution, since efficient S-protein cleavage is a requirement for successful infection, and a determinant for zoonosis or overcoming species barriers.

Removal of the PRRAR S1-S2 furin cleavage site from SARS-CoV-2 has distinct effects on pseudoviral entry into different cell types. Whereas Furin cleavage site removal improved entry into VERO cells it decreased entry into Calu-3 lung cells and did not make a difference for Caco-2 GI cells (108). Furthermore addition of PRRAR to the RaTG13 S1-S2 furin cleavage site had effects on pseudoviral entry as well, where for example it allowed for improved entry into murine ACE2 expressing cells, whereas it decreased abrogated entry into horseshoe bat and pangolin ACE2 using cells (108). This suggests a role that the furin cleavage site in SARS-CoV-2 may have contributed to the expanded cell tropism as well as the zoonotic potential that is displayed by SARS-CoV-2.

In another study removal of the polybasic S1-S2 ( $\Delta$ CS) furin cleavage site from SARS-CoV-2 virus improved replication in VERO E6 cells (a cell line used for passaging SARS-CoV-2) but impaired replication in Calu-3 (lung cell line) and human airway epithelial cells or HAE (primary cells derived from human lung tissue) (122). The SARS-CoV-2 wt virus with the S1-S2 polybasic furin cleavage had an advantage in replicating in HAE and Calu-3 cells when compared to SARS-CoV-2  $\Delta$ CS (122). Furthermore, this advantage depends on TMPRSS2, for instance replication of wt virus matched the  $\Delta$ CS when VERO E6 cells expressed TMPRSS2. TMPRSS2 appears to allow for endosomal independent entry, which can allow SARS-CoV-2 to avoid the antivirals IFITM proteins which respond to viruses in the endosome (122). In the absence of TMPRSS2 the endosomal pathway must be used for CatB/L mediated cleavage and viral entry (122). In HAE culture the  $\Delta$ CS virus replication was restored with the addition of the IFITM inhibitor AmphotB, suggesting that  $\Delta$ CS

viruses replicate poorer because they prefer using the endosomal pathway and as a result activate an antiviral response with IFITM, whereas furin cleavage site containing wt SARS-CoV-2 avoids this pathway altogether with TMPRSS2 (122). Finally, the furin cleavage site is necessary for viral transmission in ferrets. Whereas the wt SARS-CoV-2 virus was able to transmit from infected to uninfected animals, the  $\Delta$ CS virus was not able to transmit and shed to lower titers. Sequences from SARS-CoV-2 patients showed that the removal of the furin cleavage site was at very low frequency altogether suggesting that it is important as well for human-to-human transmission.

A final study showed that removing the furin cleavage site attenuated the pathogenesis of SARS-CoV-2 in both hamsters and ACE2 transgenic mice (123). In the case of hamsters the infection with the  $\Delta$ CS virus resulted in reduced weight loss and lower disease score as well as lower viral titers in the nasal and oral washes, as opposed to wt SARS-CoV-2 (123). The same was true for ACE2 transgenic mice where the infection with the  $\Delta$ CS virus resulted in reduced weight loss, as well as lower viral RNA titers in the nasal and oral washes as well as the lung when compared to wt SARS-CoV-2 (123). Furthermore the lung histopathology was less severe for the  $\Delta$ CS virus infected mice than for the wt SARS-CoV-2 infected mice (123). Finally there were lower quantities of chemokines (a marker for immunopathology) in the lungs of  $\Delta$ CS virus infected mice than for the wt SARS-CoV-2 infected mice (123). This suggests that the furin protease cleavage site may help in part to enhance the pathogenicity of SARS-CoV-2 in humans.

Within all of the sarbecoviruses SARS-CoV-2 is the only virus that has been identified to contain a polybasic cleavage site at the S1-S2 position. In other coronaviruses there are examples of furin-like-cleavage sites being acquired independently through viral evolution, with furin cleavage sites present in MERS, HCoV-OC43, and HCoV-HKU1 (97, 124). Interestingly a bat virus RmYN02 has been identified with a PAA insertion at the S1-S2 cleavage site suggest that insertions in the S1-S2 can take place in other strains (125). Altogether this suggests that other SARS-like coronaviruses could emerge with features similar to SARS-CoV-2 such as a consensus of the RBD-motif or a furin cleavage site at S1-S2, which could result in future outbreaks.

## **Innate immune responses to SARS-CoV-2**

The PRRs that recognize SARS-CoV-2 are not determined yet but are most likely TLR3, TLR7, RIG1, and MDA5, which are the main PRRs recognizing foreign RNA PAMPs, and stimulating the interferon response (27) (see section on innate immunity for influenza). The regulation of the interferon response in COVID-19 is critical in the disease outcome and is not entirely understood.

In vitro infection of SARS-CoV-2 in cell lines resulted in low induction of Type I and Type II IFNs, and consequently moderate ISG levels, and a distinct pro-inflammatory cytokine signature consisting of IL6, IL1b, and TNF as well as many chemokines (27). Type I IFNs were also at low levels in SARS-CoV-2 infected patients independent of disease severity (27). In PBMCs of infected individuals, type I IFN expression was also low, though there was a transient early burst of IFNa possibly from the lungs. In mild to moderate COVID-19 patients there was a decline of IFNa and IFNλ after the first week, whereas in severe patients there is an increase of Type I and III IFNs during the second week (27). In a murine model, it was found that Type I IFNs do not control infection but rather are determinants of pathogenesis.

SARS-CoV-2 employs multiple tools to antagonize the IFN response as a method of immune evasion, with a total of 10 proteins so far identified. Nsp16 prevents recognition of viral RNA by intracellular RNA sensors, as well as shut down cellular RNA splicing (27). Nsp1 as described prevents binding of mRNA to the ribosome shutting down translation of cellular proteins. Nsp8 and nsp9 prevent protein trafficking to the plasma membrane. Altogether the result is the suppression of type I interferon responses in the cell (27). Identification of the viral host cell interactome implicated other nsps and ORFs in the suppression of Type I IFN activation such as nsp13, ORF6, and nsp15 and ORF9b which have been confirmed to suppress Type I IFNs in functional analysis (27). Altogether SARS-CoV-2 attenuated the Type I IFN response in various different ways, which interferes with

the crosstalk between antiviral, proinflammatory and adaptive immune responses, with implications for the severity and outcome of COVID-19 (27).

Soluble factors such as cytokines, chemokines, inhibitory factors and metabolites, all work together to orchestrate the immune responses, the immune cell functions, and the immune cell localization (locally and systemically) (27). Therefore they play an important role in disease, especially when their levels and functions are dysregulated (27). This type of soluble factor dysregulation is known as the “cytokine storm” or cytokine release syndrome (CRS), and has been implicated in the severe disease outcomes that are seen in some COVID-19 patients (27).

In the lungs SARS-CoV-2 enters and rapidly replicates in type II pneumocytes resulting in a proinflammatory state where cytokines such as IL6, IL1b, and TNF are elevated. Infection of monocytes but not dendritic cells results in a similar cytokine manifestation suggesting infected monocytes may be the source of this signature (27). The cytokines involved in the NF- $\kappa$ B pathway appear to be mostly involved in this case, where NF- $\kappa$ B dependent cytokines can result in the accumulation of both neutrophils and macrophages in the lung (27). The presence of neutrophils and macrophages further induces more inflammatory cytokines and chemokines in the bronchoalveolar lavage fluid (BALF) ( CCL2, CCL3, CCL4, and CXCL10) as well as systematically through circulation (IL-1, IFNg, IL-17, TNF, IP-10, MCP-1, G-CSF, GM-CSF, IL-1RA, CCL2, CCL3, CCL5, CCL8, CXCL2, CXCL8, CXCL9, and CXCL16) (27).

These types of cytokine signatures were observed in COVID-19 patients particularly during the second week of disease in severe patients (27). These cytokine profiles in combination with the neutrophil and macrophage infiltration in the lungs result in excessive inflammation in the lung, acute respiratory distress syndrome (ARDS), characterized by pulmonary edema, lung epithelial necrosis, vascular endothelial damage, and multi-organ system failure (27). Several clinical laboratory parameters can be used as a surrogate marker for this cytokine signature including alanine aminotransferase, lactate dehydrogenase, C-reactive protein (CRP), ferritin, and D-dimers (27). IL-6 and TNF have been shown to be elevated in blood immune cells. The elevation of TNF as well as

IFNy is sufficient for the induction of pyroptosis, apoptosis, and PANoptosis (necrosis) and resulting inflammatory cell death seen in the infected lung tissue and driving COVID-19 mortality. In more mild disease elevated IFNy and IL10 are another distinct signature (27). Longitudinal studies in COVID-19 patients reveal a dynamic profile of cytokine levels. Another feature of the overexuberant activation of inflammation and cytokine storm is the failure of the inflammation to resolve (27).

In children COVID-19 disease is mostly mild but there are rare cases of a multi system inflammatory syndrome or MIS-C which manifest a couple of months after infection (27). The inflammatory cytokine profile is similar to another pediatric inflammatory disease (Kawasaki, which causes vasculitis) as well as acute severe covid disease, however it also has features that are unique. Autoantibodies might also be implicated in the pathogenesis of MIS-C (27).

The recognition of SARS-CoV-2 virus by innate immune cells drives innate immune responses which bring in more innate immune cells as well as activating the adaptive immune response for the purpose of viral clearance (27). In COVID-19 patients infected respiratory epithelial cells release chemokines that result in an influx of monocytes and neutrophils to the nasopharynges, BALF, and parts of the lung. In severe cases, chemokine levels are elevated, and neutrophils recruited to higher levels. Proinflammatory macrophages are also seen in severe patients which may lead to pathology by the recruitment of granulocytes and monocytes to the lungs (27).

In another study elevated neutrophils and NK cells, with lower amounts of T-cells and dendritic cells were seen in the BALF of severe patients, whereas clonally expanded CD8-T-cells were seen in moderately infected patients (27). The increased levels of neutrophils with respect to lymphocytes (high NLR ratio) are indeed a hallmark of severe COVID-19 disease. Another feature is the dysfunction of CD14+ monocytes potentially implicated in their ability to present antigen for activation of adaptive immune responses (27). The neutrophils in severe COVID-19 have been shown to have activated neutrophil extracellular traps (NETs) formation. NETosis can seed immunothrombosis, with platelets and neutrophils interacting and guiding thrombus formation, which

could explain the pro-thrombotic complications and coagulopathy seen in severe COVID-19 patients. Plasmacytoid dendritic cells are also reduced and impaired in function in severe COVID-19 patients (27). Lymphopenia or the reduction of lymphocytes was also seen in severe COVID-19, as well as a reduction in levels of NK-cells as well as NK-cell exhaustion. Plasmacytoid dendritic cells (pDCs) are also reduced and impaired in function in severe COVID-19 patients. Conventional dendritic cells (cDCs) are also impaired with reduced activation and therefore poorer antigen presentation for activation of the adaptive immune response (27).

### **Adaptive immune response to SARS-CoV-2: T-cell responses**

The innate immune dysregulation described above has been shown to result in aberrant T-cell responses, therefore studies probing the T-cell responses in convalescent patients have been carried out. The magnitude and composition of the CD4+ and CD8+ T-cells reactive to SARS-CoV-2 have been identified in COVID-19 patients (126). SARS-CoV-2 infection resulted in robust CD4+ and CD8+ T-cell responses in a majority of convalescent COVID-19 patients (126). For CD4+ T-cell responses a majority of the T-cell responses were elicited towards the spike followed by M and N which matched protein abundance during infection. CD4+ T-cells that recognize the ORFs and nsps were at a much lower abundance with ORF7a, ORF3a, and ORF8 being the best recognized (126). Furthermore CD4+ T-cell responses correlated with serum IgG titers against the spike RBD. CD8+T-cell responses also predominantly targeted spike, followed by M followed by N, with ORF3a and ORF8 being the best recognized from the remainder of the ORFeome (126). Interestingly, in unexposed individuals there was the presence of SARS-CoV-2 specific CD4+ and CD8+ T-cells suggesting that these are cross-reactive T-cells elicited by infections of other human coronaviruses (126). Altogether this revealed that strong T-cell responses occurred in recovered patients of COVID-19, with a clear ID hierarchy favoring the highly expressed structural proteins, with the CD4 T-cell response playing a role in improving the antibody response (126).

T-cell responses were also tracked 6 months post infection. CD4+ and CD8+ T-cell responses were around 50% higher in symptomatic vs asymptomatic individuals (127). Furthermore CD4+ T-cells correlated with antibody titers 6 months after infection (127).

It is estimated that 20-50% of uninfected individuals have CD4+ T-cells directed against SARS-CoV-2 viral proteins (128). The CD4+ T-cell repertoire was analyzed in uninfected individuals using a peptide pool that displayed peptides from different SARS-CoV-2 proteins (128). The most vigorous CD4+ T-cell responses in uninfected individuals targeted Spike protein accounting for around 50% of the response (128). The epitopes targeted by the CD4+ T-cells of uninfected individuals were cross-reactive for other HCoVs. Furthermore a majority of these cross-reactive CD4+ T-cells originated from the memory T-cell compartment suggesting that they were elicited from prior infection (128). Further inspection of these cross-reactive memory-T-cells suggested that they had comparable affinity with peptides from SARS-CoV-2 and the other 4 HCoVs, HCoV-OC43, HCoV-229E, HCoV-NL63, and HCoV-HKU1. This existence of CD4+ T-cell memory could partially explain the differences in disease outcomes in COVID-19 patients (128).

T-cells elicited by SARS-CoV-2 display acquire normal memory phenotypes suggesting that T-cell immunity can be recalled during reinfection (129). Both T-cell hyperactivation and T-cell exhaustion phenotypes have been seen in COVID-19. For instance CD38+ expression on T-cells, a marker of T-cell Activation, correlated with disease severity (129). Additionally, decreases in regulatory T-cell subsets such as T-regps have been shown as well. Markers for CD8+ T-cell exhaustion (overstimulation of T-cells that results in dampening of activity usually a feature of chronic infection) and immunosuppression have been observed following infection (129). There is evidence that T-cell memory recall responses are not impaired after recovery (129). There is a wide variety of different conflicting phenotypes that have been reported for T-cells, suggesting that the T-cell responses to COVID-19 are highly variable owing to the variability seen in disease severity.

Studies trying to characterize the immunotype profiles, or the characteristics of the immune cell types, kinetics and activity, were carried out for patients that represented mild, moderate, and severe COVID-19 disease (130). There were 3 immunotypes observed that featured distinct T-cell phenotypes. Immunotype 1 featured robust CD4+ T-cell responses, highly activated or exhausted CD8+ T-cells, with poorer  $T_{FH}$  responses (CD4+ T-follicular helper cells) (130). Immunotype 2 instead had less robust CD4+ T-cell responses, with effector like CD8+ T-cells, and plasmablast and MBC responses, Immunotype 3 had lack of lymphocyte responses that were almost indistinguishable from uninfected individuals (130). Immunotype 1 correlated with the most severe disease which matches the overexuberant, exaggerated, and dysregulated immune responses (130). Other signatures associated with disease severity were identified, though it is hard to exactly identify the determinants and mechanisms of disease severity using these studies (129).

$T_{H1}$  CD4+ T-cell responses with balanced  $T_{H1}/T_{H2}$  and CTL responses are generally hallmarks of T-cell mediated immunity against viral infections (129). In COVID-19 convalescent individuals  $T_{H1}$  CD4+ T-cells are dominantly elicited with IFN-g, TNF-a, and IL-2 cytokine expression profiles, which were present in both mild and severe cases (129). Additionally CD8+ T-cells featuring IFN-g, TNF-a, GzmB, and/or CD107a, which are markers of CD8+ T-cell cytotoxicity and degranulation) have been identified on convalescent patients, though increases of these cell types have correlated with more severe outcomes (129). In some but not cases of severe disease,  $T_{H2}$  and  $T_{H3}$  responses have been identified which may be associated with pro-inflammatory cytokine secretion that result in neutrophilia, excess inflammation, tissue and organ damage as well as impaired T-cell function (129). It is also possible that T-cell recognition of self-epitopes or superantigens during COVID-19 infection may result in the tissue damage seen in severe cases as well as the MIS-C syndrome in children (129).

$T_{FH}$  responses are required for the generation of affinity matured neutralizing antibodies. In COVID-19 convalescent individuals who have mild or moderate disease, The S-specific  $T_{FH}$  re-

sponses were characterized by  $T_H1/2$  and  $T_H17$ , with  $T_H1/2$  correlating the best with plasma neutralization potency (129). Furthermore,  $T_{FH}$  cells with germinal center activity signatures have been identified in convalescent individuals. The proportion of  $T_H17$  and the total S-specific  $T_{FH}$  inversely correlate with disease severity where  $T_H17$  skewed responses generally appear in more severe cases (129).

### **Adaptive immune response to SARS-CoV-2: B-cell responses and antibodies**

The B-cell and humoral antibody response to COVID-19 is currently thought to be one of the main correlates of protection against SARS-CoV-2 in both vaccines and against reinfection (131). A majority of the antibodies detected in SARS-CoV-2 infection target the two structural proteins S and N (132). A major focus of the antibody response against SARS-CoV-2 is to the S protein particularly the RBD since it mediates viral entry and is a target of neutralizing antibodies (133). In convalescent COVID-19 patients serum antibodies that were screened for binding of S protein and RBD via ELISA revealed highly heterogeneous responses. Serum antibody levels appeared to wane over time but were relatively stable across all age groups. The neutralizing antibody response against SARS-CoV-2 S pseudotyped lentivirus as well as authentic SARS-CoV-2 also matched this high heterogeneity with some individuals having high serum neutralization ( $>5000$  ID<sub>50</sub>) whereas a majority featured low serum neutralization ( $<50$  ID<sub>50</sub>) (133). Hospitalized patients appeared to have a modest significantly higher serum neutralization and binding to S protein. Binding to S protein as well as RBD did correlate with serum neutralization potency (133). Although serum neutralization was not particularly high in most individuals, nonetheless antibody sequencing from B-cells was able to pull out highly neutralizing antibodies that bound the RBD (133).

The polyclonal antibody response to SARS-CoV-2 in convalescent COVID-19 individuals was characterized by ELISAs and polyclonal EM of plasma IgGs to identify binding features of anti-S anti-

bodies (134). Binding of plasma IgGs to SARS-CoV-2 S protein and RBD was strongest as expected. Binding of plasma IgGs against SARS-CoV revealed some IgG binding to full S protein as well as RBD, whereas binding to MERS was limited to full S protein only (134). This makes sense as there is some sequence conservation between SARS-CoV-2 S protein and S-RBD whereas for MERS-CoV the S2 domain is the only domain that is well conserved. There were plasma IgG responses to the S protein and RBD to other HCoV alpha and beta coronaviruses, but these were most likely elicited from prior in (134). Polyclonal EM revealed that antibodies targeted both the RBD as well as the NTD, suggesting that these domains may be immunodominant. Further characterization of monoclonals from these individuals revealed that potently neutralizing antibodies that targeted the RBD, occluded the ACE-2 binding site on the RBD therefore potentially blocking ACE2 binding (134).

Since the RBD of Spike is immunodominant, the distinct epitopes on the spike protein were further characterized through structural studies of eight monoclonals isolated from convalescent patients that targeted the RBD (135). The structures solved using both single particle cryo-EM and X-ray Crystallography revealed 4 distinct classes of antibodies that recognize 4 different epitopes (135). Class 1 antibodies recognize the ACE2 binding motif (hence block ACE2 binding) and only recognize RBDs that are in the up conformation. Class 2 antibodies, which recognizes both up and down conformations, blocks the ACE2 binding motif as well, and can contact adjacent RBDs (135). Class 3 antibodies that bind outside of the ACE2 binding motif and recognize both up and down conformations. Class 4 antibodies that do not bind the ACE2 binding site and only recognize the up conformation. Interestingly somatic hypermutation of these antibodies did not map to the antibody paratopes, suggesting that these binding capabilities are completely germline encoded (135). Most of these antibody neutralizing functions depend on their ability to block the interaction of the RBD with ACE2 (135). Of note is that class 4 epitopes are highly conserved between sarbecoviruses and therefore class 4 antibodies are highly broad, although because they do not block ACE2 binding as well, they are less potent (135–137). Class 3 and well as class 4 antibodies were isolated from

SARS-CoV convalescent individuals being able to cross react with and neutralize SARS-CoV-2 (135–137).

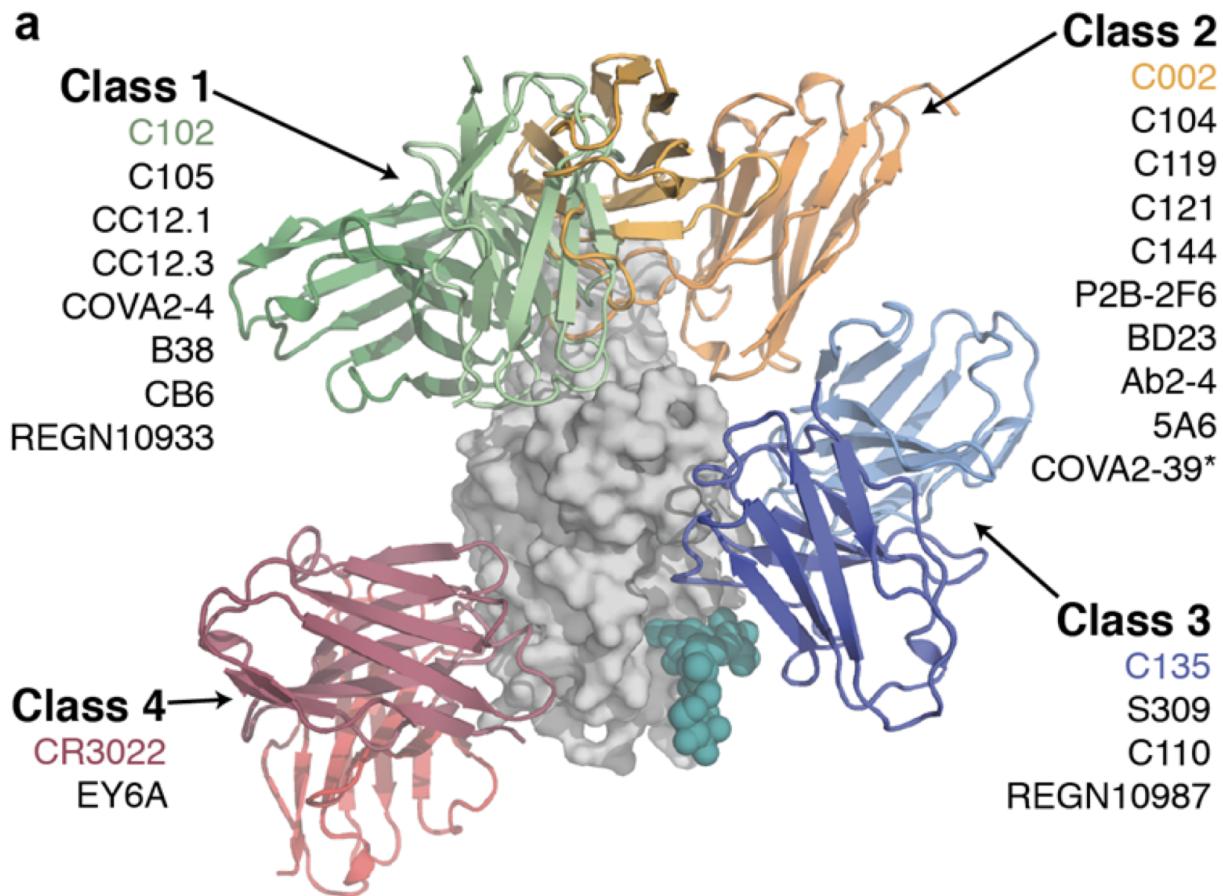


Figure 9: Structural depiction of 4 different classes of antibodies that recognize the RBD. Taken from (135).

Longitudinal studies that tracked the antibody responses as well as B-cell responses overtime have also been carried out. In both mild and severe convalescent patients although serum neutralizing titers waned over the period of 5 months, S-reactive IgG MBCs were stable or increased. Isolating monoclonals at different timepoints from these individuals revealed that over time the somatic hypermutation increased, as well as binding affinity and neutralization potency. The B-cell ID hierarchy was also characterized at different time points via binding of the monoclonal panel to S-RBD, ACE2

blocking vs non-ACE2 blocking, NTD, other S1 and S2. Broad coverage of these epitopes was seen in most individuals and the ID hierarchies remained stable over time. Interestingly quite a large percent of monoclonals recognized the S2 domain. A relatively low percentage of antibodies were potently neutralized with the large majority recognizing the ACE-binding site of RBD, and only a small fraction recognizing non-ACE2 RBD. The rest of the epitopes were largely non-neutralizing. Neutralization potency did however increase over time. Finally, only the S2 binding monoclonals were cross reactive to HCoV-OC43 and HCoV-HKU1, and these monoclonals started out with a high level of SHM, suggesting that these antibodies were activated by a recall response to HCoVs. On the other hand the antibodies cross-reactive to SARS-CoV recognized both RBD as well as S2, and these antibodies started out with a similar SHM to SARS-CoV-2 specific antibodies suggesting that these antibodies were elicited from the naive response to SARS-CoV-2 infection (138). These antibodies that are cross-reactive between SARS-CoV-2 and SARS-CoV may be class 3 and class 4 like antibodies.

Another study looked at longitudinal analysis 1.3 months and 6 months after infection. The study confirmed IgG titer and neutralizing titer waning with the concomitant maintenance or increase of S-reactive MBCs, and higher levels of SHM 6 months after infection (139). Interestingly the clonal diversity changed after 6 months with fewer clones expanded after 1.3 months present in the 6 months MBC repertoire, and low overlap between the two timepoints. This suggests that there may be prolonged GC interactions with extensive clonal turnover and antibody evolution, also suggesting the prolonged presence of antigens in GCs (139). Isolating monoclonals from both timepoints indicated that although the overall affinity against the RBD between the two timepoints was not significantly different, the affinity between paired antibodies from the same lineage isolated at two time points was significant with the 6-month antibodies showing higher affinity (139). The antibodies from the 6-month time point were also more resistant to RBD mutants. These binding features matched neutralization where the 6-month paired antibodies were more potent than the 1.3-month antibodies against wt as well as the mutant SARS-CoV-2 viruses. Finally antigen persistence was

seen in the bowel of asymptomatic individuals at 4 months identified by immunofluorescence staining for NP as well as PCR of intestinal biopsies (139). Altogether this study suggests a model of antigen persistence after SARS-CoV-2 infection that drives prolonged GC reactions characterized by high clonal turnover as well as antibody evolution that produces antibodies with higher binding and neutralization as well as resistance to variants of concern (139).

In a study that tracking protection mediated by anti-SARS-CoV-2 humoral immunity revealed that IgG+ titers against SARS-CoV-2 were not substantially different between deceased and recovered individuals (140). However a higher proportion of the humoral response to S protein improved outcomes and was enriched in recovered individuals whereas an anti-NP response that was skewed towards ADCC and other functional responses correlated with deceased patients (140). The immunodominant recognition of spike in convalescent individuals correlated with better outcomes in a larger set of individuals with skewed responses towards NP correlating with deceased patients (140). Another study confirmed this result where a large cohort study revealed that a larger proportion of anti-S or anti-RBD antibodies compared to anti-NP were seen in mild cases versus severely ill or deceased individuals (132). Furthermore higher IgG titers to RBD and S correlated with decreases in RNAemia, but did not necessarily correlate with disease outcome (132). Altogether these studies suggest that the quality and type of the antibody responses against SARS-CoV-2 rather than the amount influence disease outcomes.

## **SARS-CoV-2 Vaccines**

Perhaps the greatest achievement in modern science is the striking efficacy and effectiveness of the SARS-CoV-2 vaccines. To date a total of three vaccines have been approved for emergency use authorization in the United States which include the Moderna and Pfizer mRNA vaccines as well as the Johnson and Johnson Adenovirus vectored vaccines. Currently there are 5 major types of vaccines in the advanced stages of clinical testing. This includes mRNA vaccine, replication

defective adenovirus vectored vaccines, inactivated pathogen vaccines, protein subunit vaccines, and VLP vaccines (141).

mRNA vaccines are composed of a viral mRNA encapsulated in a proprietary lipid nanoparticle, and have been shown to induce robust humoral and cellular immunity (142). When injected intramuscularly they can be taken up by APCs such as dendritic cells or macrophages. Since they are composed of RNA, they can be recognized by the cytosolic RNA sensors activating innate immune responses (142). Furthermore the lipid carrier can also be recognized as a PAMP, therefore mRNA vaccines are not adjuvanted (142). In APCs they can also be translated and used to present peptides to both MHC1 and MHC2 inducing strong CD4+ and CD8+ T-cell responses (142). Finally they can be taken up by cells to produce viral proteins that are recognized by the B-cell compartment and result in a strong humoral response (142). One disadvantage of mRNA vaccines is that they require long-term cold storage (-70°C for Pfizer or -20°C for Moderna) and can only be stored at 4°C for short periods of time, due to the unstable nature of RNA.

Both the Pfizer BNT162b2 and Moderna mRNA-1273 vaccines contain an mRNA that encodes for a S trimer that is stabilized in the prefusion conformation, and the immunization regime is 2 doses (143, 144). Both of these vaccines elicited robust antibody responses as well as T-cell responses in a two dose regimen (143, 144). Finally, both vaccines were shown to be safe and efficacious against disease (94.1% for Moderna mRNA-1273, and 95% for Pfizer BNT162b2). Finally use of both vaccines in the US were shown to be effective at preventing infection as well in health care workers with 90% protection for full immunization and 80% for partial immunization. Similar reports out of Israel for the Pfizer BNT162b2 showed real life reduction in COVID-19 disease incidence and severity demonstrating the robust protection of these mRNA vaccines.

A study of the B-cell responses elicited on SARS-CoV-2 mRNA vaccination in naive and recovered individuals revealed distinct MBC and antibody responses (131). For naive individuals prime and boost of the mRNA vaccines were required for most optimal antibody responses deter-

mined by ELISA serum binding and neutralization (131). This was especially true for variant neutralization where the neutralization for naive individuals after the prime was only detectable against the B.1.351 after the boost, however the neutralization potency against B.1.351 was significantly lower than for D614G (131). For recovered individuals antibody responses peaked after the prime with no change boost for both serum binding and neutralization, including the neutralization against the variants, with equivalent neutralization titers against D614G and B.1.351 variants (131). MBC reactive to the SARS-CoV-2 spike were efficiently elicited after prime and reached maximum levels after boosting in naive individuals after vaccination. Similar to the antibody levels elicited in the SARS-CoV-2 recovered cohort, MBC levels peaked after the prime but did not change after the boost. Furthermore, MBC levels elicited after the prime in SARS-CoV-2 exposed individuals correlated with the pre-vaccination MBC levels, suggesting a successful recall of memory in SARS-CoV-2 infected individuals (131).

Adenovirus vectored vaccines use a replication incompetent adenovirus as a carrier for viral genes encoded as DNA. They have similar advantages to RNA vaccines where in this case the DNA can be recognized by cytosolic DNA sensors, with efficient presentation of peptides to both MHC-1 and MHC-2, and expression of the viral genes for humoral antibody responses (141). They are also not adjuvanted and are stable long-term at more reasonable temperatures (4°C) (141). There are four Ad-vectored vaccines currently in use globally: the Jansen JNJ-78436735/Ad26.COV2 from Jansen a Johnson & Johnson subsidiary, AZD1222 (ChAdOx1) from AstraZeneca/Oxford University, Gam-COVID-Vac/Sputnik V from Gamaleya Research Institute/Russian institute of health, and the Ad5-nCoV from CanSino Biological. All four vaccines use different adenoviruses as the vaccine vector, Human Ad26 for Jansen NJ-78436735, Chimpanzee ChAdOx1 for AstraZeneca AZD1222, Human Ad26 (for prime) and Ad5 (for boost) for Gam-COVID-Vac/Sputnik V, and Human Ad5 for Ad5-nCoV from CanSino Biological (141). Only the JNJ Ad26.COV2 vaccine uses a perfusion stabilized S protein. Remarkably the efficacy (prevention of moderate and severe disease) for the single dose regimen of the JNJ Ad26.COV2 vaccine reached

76.7%, whereas in South Africa where the B.1.351 variant was circulating it reached 64.0% (145). This showed that the Ad-vectored vaccines could be efficacious with a single dose. Some of the other Ad vectored vaccines. The Gam-COVID-Vac/Sputnik V was also highly efficacious where the Ad26 prime and an Ad5 boost regimen resulted in 96.1% efficacy (146).

Altogether vaccines have been shown to be both safe and effective in clinical trials as well as in real-world scenarios. It is unclear whether these vaccines will provide cross-protection to other coronavirus including ones that may emerge in the future. Three coronaviruses emerged within the last two decades. Furthermore, there is a large viral reservoir that exists within bats and other mammalian species globally. Therefore, it is likely that these coronaviruses will be the source for future outbreaks. This emphasizes the need for strategies to combat these viral threats to human health, with one potential solution being the development of a universal coronavirus vaccine that would protect against a wide swath of coronaviruses from current and future coronavirus strains. In Chapter 3, a mosaic nanoparticle immunization strategy will be described that elicits a polyclonal antibody response with high cross-reactivity and cross-neutralization potency to all viruses in the sarbecovirus lineage, suggesting a promising route in the development of a pan-coronavirus vaccine.

### **Bibliography:**

1. J. K. Taubenberger, D. M. Morens, 1918 Influenza: the Mother of All Pandemics. *Emerg Infect Dis.* **12**, 15–22 (2006).
2. A. Gagnon, M. S. Miller, S. A. Hallman, R. Bourbeau, D. A. Herring, D. JD. Earn, J. Madrenas, Age-Specific Mortality During the 1918 Influenza Pandemic: Unravelling the Mystery of High Young Adult Mortality. *Plos One.* **8**, e69586 (2013).
3. H. Markel, H. B. Lipman, J. A. Navarro, A. Sloan, J. R. Michalsen, A. M. Stern, M. S. Cetron, Nonpharmaceutical Interventions Implemented by US Cities During the 1918-1919 Influenza Pandemic. *Jama.* **298**, 644–654 (2007).

4. T. Watanabe, Y. Kawaoka, Pathogenesis of the 1918 Pandemic Influenza Virus. *Plos Pathog.* **7**, e1001218 (2011).
5. D. M. Morens, A. S. Fauci, The 1918 Influenza Pandemic: Insights for the 21st Century. *J Infect Dis.* **195**, 1018–1028 (2007).
6. A. H. Reid, T. G. Fanning, J. V. Hultin, J. K. Taubenberger, Origin and evolution of the 1918 “Spanish” influenza virus hemagglutinin gene. *Proc National Acad Sci.* **96**, 1651–1656 (1999).
7. M. Worobey, G.-Z. Han, A. Rambaut, Genesis and pathogenesis of the 1918 pandemic H1N1 influenza A virus. *Proc National Acad Sci.* **111**, 8107–8112 (2014).
8. F. Krammer, G. J. D. Smith, R. A. M. Fouchier, M. Peiris, K. Kedzierska, P. C. Doherty, P. Palese, M. L. Shaw, J. Treanor, R. G. Webster, A. G. x000ED a-Sastre, Influenza. *Nature Reviews Disease Primers.* **4**, 1–21 (2018).
9. S. Herfst, E. J. A. Schrauwen, M. Linster, S. Chutinimitkul, E. de Wit, V. J. Munster, E. M. Sorrell, T. M. Bestebroer, D. F. Burke, D. J. Smith, G. F. Rimmelzwaan, A. D. M. E. Osterhaus, R. A. M. Fouchier, Airborne transmission of influenza A/H5N1 virus between ferrets. *Science.* **336**, 1534–1541 (2012).
10. A. D. Iuliano, K. M. Roguski, H. H. Chang, D. J. Muscatello, R. Palekar, S. Tempia, C. Cohen, J. M. Gran, D. Schanzer, B. J. Cowling, P. Wu, J. Kyncl, L. W. Ang, M. Park, M. Redlberger-Fritz, H. Yu, L. Espenhain, A. Krishnan, G. Emukule, L. van Asten, S. P. da Silva, S. Aungkulanon, U. Buchholz, M.-A. Widdowson, J. S. Bresee, G. S. I. M. C. Network, E. Azziz-Baumgartner, P.-Y. Cheng, F. Dawood, I. Foppa, S. Olsen, M. Haber, C. Jeffers, C. R. MacIntyre, A. T. Newall, J. G. Wood, M. Kundi, T. Popow-Kraupp, M. Ahmed, M. Rahman, F. Marinho, C. V. S. Proschle, N. V. Mallegas, F. Luzhao, L. Sa, J. Barbosa-Ramírez, D. M. Sanchez, L. A. Gomez, X. B. Vargas, aBetsy A. Herrera, M. J. Llanés, T. K. Fischer, T. G. Krause, K. Mølbak, J. Nielsen, R. Trebbien, A. Bruno, J. Ojeda, H. Ramos, M. an der Heiden, L. del C. C. Signor, C. E. Serrano, R. Bhardwaj, M. Chadha, V. Narayan, S. Kosen, M. Bromberg, A. Glatman-Freedman, Z. Kaufman, Y. Arima,

K. Oishi, S. Chaves, B. Nyawanda, R. A. Al-Jarallah, P. A. Kuri-Morales, C. R. Matus, M. E. J. Corona, A. Burmaa, O. Darmaa, M. Obtel, I. Cherkoui, C. C. van den Wijngaard, W. van der Hoek, M. Baker, D. Bandaranayake, A. Bissielo, S. Huang, L. Lopez, C. Newbern, E. Flem, G. M. Grøneng, S. Hauge, F. G. de Cosío, Y. de Moltó, L. M. Castillo, M. A. Cabello, M. von Horoch, J. M. Osis, A. Machado, B. Nunes, A. P. Rodrigues, E. Rodrigues, C. Calomfirescu, E. Lupulescu, R. Popescu, O. Popovici, D. Bogdanovic, M. Kostic, K. Lazarevic, Z. Milosevic, B. Tiodorovic, M. Chen, J. Cutter, V. Lee, R. Lin, S. Ma, A. L. Cohen, F. Treurnicht, W. J. Kim, C. Delgado-Sanz, S. de mateo Ontañón, A. Larrauri, I. L. León, F. Vallejo, R. Born, C. Junker, D. Koch, J.-H. Chuang, W.-T. Huang, H.-W. Kuo, Y.-C. Tsai, K. Bundhamcharoen, M. Chittaganpitch, H. K. Green, R. Pebody, N. Goñi, H. Chiparelli, L. Brammer, D. Mustaquim, Estimates of global seasonal influenza-associated respiratory mortality: a modelling study. *Lancet.* **391**, 1285–1300 (2018).

11. S. A. Cohen, K. K. H. Chui, E. N. Naumova, Influenza vaccination in young children reduces influenza-associated hospitalizations in older adults, 2002-2006. *J Am Geriatr Soc.* **59**, 327–32 (2011).

12. D. Mertz, T. H. Kim, J. Johnstone, P.-P. Lam, M. Science, S. P. Kuster, S. A. Fadel, D. Tran, E. Fernandez, N. Bhatnagar, M. Loeb, Populations at risk for severe or complicated influenza illness: systematic review and meta-analysis. *Bmj Br Medical J.* **347**, f5061 (2013).

13. E. Böttcher-Friebertshäuser, H. Klenk, W. Garten, Activation of influenza viruses by proteases from host cells and bacteria in the human airway epithelium. *Pathog Dis.* **69**, 87–100 (2013).

14. J. R. Tisoncik, M. J. Korth, C. P. Simmons, J. Farrar, T. R. Martin, M. G. Katze, Into the Eye of the Cytokine Storm. *Microbiology and Molecular Biology Reviews.* **76**, 16–32 (2012).

15. M. Moriyama, W. J. Hugentobler, A. Iwasaki, *Annual Review of Virology*, in press, doi:10.1146/annurev-virology-012420-022445.

16. D. J. Alexander, An overview of the epidemiology of avian influenza. *Vaccine.* **25**, 5637–5644 (2007).

17. WHO, Avian Influenza Weekly Update Number 7

8

8 (2021), (available at

18. WHO, Cumulative number of confirmed human cases for avian influenza A(H5N1) reported to WHO, 2003-2020 (2021), (available at [https://www.who.int/influenza/human\\_animal\\_interface/2020\\_DEC\\_tableH5N1.pdf?ua=1](https://www.who.int/influenza/human_animal_interface/2020_DEC_tableH5N1.pdf?ua=1)).

19. R. Andino, E. Domingo, Viral quasispecies. *Virology*. **479**, 46–51 (2015).

20. N. Tzarum, R. P. de Vries, X. Zhu, W. Yu, R. McBride, J. C. Paulson, I. A. Wilson, Structure and Receptor Binding of the Hemagglutinin from a Human H6N1 Influenza Virus. *Cell Host Microbe*. **17**, 369–376 (2015).

21. C. M. Mair, K. Ludwig, A. Herrmann, C. Sieben, Receptor binding and pH stability — How influenza A virus hemagglutinin affects host-specific virus infection. *Biochimica Et Biophysica Acta Bba - Biomembr*. **1838**, 1153–1168 (2014).

22. H. Zhang, X. Li, J. Guo, L. Li, C. Chang, Y. Li, C. Bian, K. Xu, H. Chen, B. Sun, The PB2 E627K mutation contributes to the high polymerase activity and enhanced replication of H7N9 influenza virus. *J Gen Virol*. **95**, 779–786 (2014).

23. É. A. Moreira, S. Locher, L. Kolesnikova, H. Bolte, T. Aydillo, A. García-Sastre, M. Schwemmle, G. Zimmer, Synthetically derived bat influenza A-like viruses reveal a cell type- but not species-specific tropism. *Proc National Acad Sci*. **113**, 12797–12802 (2016).

24. U. Karakus, T. Thamamongood, K. Ciminski, W. Ran, S. C. Günther, M. O. Pohl, D. Eletto, C. Jeney, D. Hoffmann, S. Reiche, J. Schinköthe, R. Ulrich, J. Wiener, M. G. B. Hayes, M. W. Chang, A. Hunziker, E. Yángüez, T. Aydillo, F. Krammer, J. Oderbolz, M. Meier, A. Oxenius, A. Halenius, G. Zimmer, C. Benner, B. G. Hale, A. García-Sastre, M. Beer, M. Schwemmle, S. Stertz, MHC class II proteins mediate cross-species entry of bat influenza viruses. *Nature*. **567**, 1–

22 (2019).

25. S. J. Zost, K. Parkhouse, M. E. Gumiña, K. Kim, S. D. Perez, P. C. Wilson, J. J. Treanor, A. J. Sant, S. Cobey, S. E. Hensley, Contemporary H3N2 influenza viruses have a glycosylation site that alters binding of antibodies elicited by egg-adapted vaccine strains. *Proc National Acad Sci.* **114**, 12578–12583 (2017).

26. M. J. Killip, E. Fodor, R. E. Randall, Influenza virus activation of the interferon system. *Virus Res.* **209**, 11–22 (2015).

27. J. L. Schultze, A. C. Aschenbrenner, COVID-19 and the human innate immune system. *Cell.* **184**, 1671–1692 (2021).

28. T. D. Cline, D. Beck, E. Bianchini, Influenza virus replication in macrophages: balancing protection and pathogenesis. *J Gen Virol.* **98**, 2401–2412 (2017).

29. A. Smed-Sörensen, C. Chalouni, B. Chatterjee, L. Cohn, P. Blattmann, N. Nakamura, L. Delamarre, I. Mellman, Influenza A Virus Infection of Human Primary Dendritic Cells Impairs Their Ability to Cross-Present Antigen to CD8 T Cells. *Plos Pathog.* **8**, e1002572 (2012).

30. Q. Liu, Y. Zhou, Z. Yang, The cytokine storm of severe influenza and development of immunomodulatory therapy. *Cell Mol Immunol.* **13**, 3–10 (2016).

31. L. Zhu, L. Liu, Y. Zhang, L. Pu, J. Liu, X. Li, Z. Chen, Y. Hao, B. Wang, J. Han, G. Li, S. Liang, H. Xiong, H. Zheng, A. Li, J. Xu, H. Zeng, High Level of Neutrophil Extracellular Traps Correlates With Poor Prognosis of Severe Influenza A Infection. *J Infect Dis.* **217**, 428–437 (2018).

32. A. J. Sant, A. T. DiPiazza, J. L. Nayak, A. Rattan, K. A. Richards, CD4 T cells in protection from influenza virus: Viral antigen specificity and functional potential. *Immunol Rev.* **284**, 91–105 (2018).

33. Z. Wang, Y. Wan, C. Qiu, S. Quiñones-Parra, Z. Zhu, L. Loh, D. Tian, Y. Ren, Y. Hu, X. Zhang, P. G. Thomas, M. Inouye, P. C. Doherty, K. Kedzierska, J. Xu, Recovery from severe H7N9 disease is associated with diverse response mechanisms dominated by CD8+ T cells. *Nat Commun.* **6**, 6833 (2015).

34. F. Krammer, The human antibody response to influenza A virus infection and vaccination. *Nature Reviews Immunology*. **19**, 1–15 (2019).

35. D. Angeletti, J. S. Gibbs, M. Angel, I. Kosik, H. D. Hickman, G. M. Frank, S. R. Das, A. K. Wheatley, M. Prabhakaran, D. J. Leggat, A. B. McDermott, J. W. Yewdell, Defining B cell immunodominance to viruses. *Nature Immunology*. **18**, 456–463 (2017).

36. A. G. Schmidt, M. D. Therkelsen, S. Stewart, T. B. Kepler, H.-X. Liao, M. A. Moody, B. F. Haynes, S. C. Harrison, Viral Receptor-Binding Site Antibodies with Diverse Germline Origins. *Cell*. **161**, 1026–1034 (2015).

37. K. R. McCarthy, A. Watanabe, M. Kuraoka, K. T. Do, C. E. McGee, G. D. Sempowski, T. B. Kepler, A. G. Schmidt, G. Kelsoe, S. C. Harrison, Memory B Cells that Cross-React with Group 1 and Group 2 Influenza A Viruses Are Abundant in Adult Human Repertoires. *Immunity*. **48**, 174–183.e9 (2018).

38. R. Nachbagauer, A. Choi, A. Hirsh, I. Margine, S. Iida, A. Barrera, M. Ferres, R. A. Albrecht, A. García-Sastre, N. M. Bouvier, K. Ito, R. A. Medina, P. Palese, F. Krammer, Defining the antibody cross-reactome directed against the influenza virus surface glycoproteins. *Nat Immunol*. **18**, 464–473 (2017).

39. D. Corti, J. Voss, S. J. Gamblin, G. Codoni, A. Macagno, D. Jarrossay, S. G. Vachieri, D. Pinna, A. Minola, F. Vanzetta, C. Silacci, B. M. Fernandez-Rodriguez, G. Agatic, S. Bianchi, I. Giacchetto-Sasselli, L. Calder, F. Sallusto, P. Collins, L. F. Haire, N. Temperton, J. P. M. Lange-dijk, J. J. Skehel, A. Lanzavecchia, A Neutralizing Antibody Selected from Plasma Cells That Binds to Group 1 and Group 2 Influenza A Hemagglutinins. *Science*. **333**, 850–856 (2011).

40. S. Ng, R. Nachbagauer, A. Balmaseda, D. Stadlbauer, S. Ojeda, M. Patel, A. Rajabhathor, R. Lopez, A. F. Guglia, N. Sanchez, F. Amanat, L. Gresh, G. Kuan, F. Krammer, A. Gordon, Novel correlates of protection against pandemic H1N1 influenza A virus infection. *Nat Med*. **25**, 962–967 (2019).

41. D. J. DiLillo, G. S. Tan, P. Palese, J. V. Ravetch, Broadly neutralizing hemagglutinin stalk–

specific antibodies require Fc $\gamma$ R interactions for protection against influenza virus in vivo. *Nat Med.* 20, 143–151 (2014).

42. H. M. Yassine, P. M. McTamney, J. C. Boyington, T. J. Ruckwardt, M. C. Crank, M. K. Smatti, J. E. Ledgerwood, B. S. Graham, Use of Hemagglutinin Stem Probes Demonstrate Prevalence of Broadly Reactive Group 1 Influenza Antibodies in Human Sera. *Nature Publishing Group.* **8**, 1–11 (2018).

43. N. Pica, R. Hai, F. Krammer, T. T. Wang, J. Maamary, D. Eggink, G. S. Tan, J. C. Krause, T. Moran, C. R. Stein, D. Banach, J. Wrammert, R. B. Belshe, A. García-Sastre, P. Palese, Hemagglutinin stalk antibodies elicited by the 2009 pandemic influenza virus as a mechanism for the extinction of seasonal H1N1 viruses. *Proc National Acad Sci.* **109**, 2573–2578 (2012).

44. R. Nachbagauer, W.-C. Liu, A. Choi, T. J. Wohlbold, T. Atlas, M. Rajendran, A. Solórzano, F. Berlanda-Scorza, A. García-Sastre, P. Palese, R. A. Albrecht, F. Krammer, A universal influenza virus vaccine candidate confers protection against pandemic H1N1 infection in preclinical ferret studies. *npj Vaccines.* **2**, 1–12 (2017).

45. A. Impagliazzo, F. Milder, H. Kuipers, M. V. Wagner, X. Zhu, R. M. B. Hoffman, R. van Meersbergen, J. Huizingh, P. Wanningen, J. Verspuij, M. de Man, Z. Ding, A. Apetri, B. Kükrer, E. Sneekes-Vriese, D. Tomkiewicz, N. S. Laursen, P. S. Lee, A. Zakrzewska, L. Dekking, J. Tolboom, L. Tettero, S. van Meerten, W. Yu, W. Koudstaal, J. Goudsmit, A. B. Ward, W. Meijberg, I. A. Wilson, K. Radošević, A stable trimeric influenza hemagglutinin stem as a broadly protective immunogen. *Science.* **349**, 1301–1306 (2015).

46. R. Nachbagauer, J. Feser, A. Naficy, D. I. Bernstein, J. Guptill, E. B. Walter, F. Berlanda-Scorza, D. Stadlbauer, P. C. Wilson, T. Aydillo, M. A. Behzadi, D. Bhavsar, C. Bliss, C. Capuano, J. M. Carreño, V. Chromikova, C. Claeys, L. Coughlan, A. W. Freyn, C. Gast, A. Javier, K. Jiang, C. Mariottini, M. McMahon, M. McNeal, A. Solórzano, S. Strohmeier, W. Sun, M. V. der Wielen, B. L. Innis, A. García-Sastre, P. Palese, F. Krammer, 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*, 1–9 (2020).

47. G. A. Sautto, G. A. Kirchenbaum, T. M. Ross, Towards a universal influenza vaccine: different approaches for one goal. *Virology Journal*. **15**, 1–12 (2018).

48. H. M. Yassine, J. C. Boyington, P. M. McTamney, C.-J. Wei, M. Kanekiyo, W.-P. Kong, J. R. Gallagher, L. Wang, Y. Zhang, M. G. Joyce, D. Lingwood, S. M. Moin, H. Andersen, Y. Okuno, S. S. Rao, A. K. Harris, P. D. Kwong, J. R. Mascola, G. J. Nabel, B. S. Graham, Hemagglutinin-stem nanoparticles generate heterosubtypic influenza protection. *Nature Medicine*. **21**, 1065–1070 (2015).

49. X. Yu, T. Tsibane, P. A. McGraw, F. S. House, C. J. Keefer, M. D. Hicar, T. M. Tumpey, C. Pappas, L. A. Perrone, O. Martinez, J. Stevens, I. A. Wilson, P. V. Aguilar, E. L. Altschuler, C. F. Basler, J. E. C. Jr, Neutralizing antibodies derived from the B cells of 1918 influenza pandemic survivors. *Nature*. **455**, 532–536 (2008).

50. D. N. Fisman, R. Savage, J. Gubbay, C. Achonu, H. Akwar, D. J. Farrell, N. S. Crowcroft, P. Jackson, Older Age and a Reduced Likelihood of 2009 H1N1 Virus Infection. *New Engl J Medicine*. **361**, 2000–2001 (2009).

51. C. W. Davis, K. J. L. Jackson, M. M. McCausland, J. Darce, C. Chang, S. L. Linderman, C. Chennareddy, R. Gerkin, S. J. Brown, J. Wrammert, A. K. Mehta, W. C. Cheung, S. D. Boyd, E. K. Waller, R. Ahmed, Influenza vaccine–induced human bone marrow plasma cells decline within a year after vaccination. *Science*, eaaz8432 (2020).

52. M. O. Altman, J. R. Bennink, J. W. Yewdell, B. R. Herrin, Lamprey VLRB response to influenza virus supports universal rules of immunogenicity and antigenicity. *Elife*. **4**, e07467 (2015).

53. E. Kirkpatrick, X. Qiu, P. C. Wilson, J. Bahl, F. Krammer, The influenza virus hemagglutinin head evolves faster than the stalk domain. *Sci Rep-uk*. **8**, 10432 (2018).

54. N. C. Wu, I. A. Wilson, A Perspective on the Structural and Functional Constraints for Immune Evasion: Insights from Influenza Virus. *J Mol Biol*. **429**, 2694–2709 (2017).

55. S. K. Dougan, J. Ashour, R. A. Karssemeijer, M. W. Popp, A. M. Avalos, M. Barisa, A. F. Altenburg, J. R. Ingram, J. J. Cragnolini, C. Guo, F. W. Alt, R. Jaenisch, H. L. Ploegh, Antigen-specific B-cell receptor sensitizes B cells to infection by influenza virus. *Nature*. **503**, 406–409 (2013).

56. F. Broecker, S. T. H. Liu, W. Sun, F. Krammer, V. Simon, P. Palese, Immunodominance of Antigenic Site B in the Hemagglutinin of the Current H3N2 Influenza Virus in Humans and Mice. *J Virol*. **92**, e01100-18 (2018).

57. A. Amitai, M. Sangesland, R. M. Barnes, D. Rohrer, N. Lonberg, D. Lingwood, A. K. Chakraborty, Defining and Manipulating B Cell Immunodominance Hierarchies to Elicit Broadly Neutralizing Antibody Responses against Influenza Virus. *Cell Syst*. **11**, 573-588.e9 (2020).

58. S. F. Andrews, Y. Huang, K. Kaur, L. I. Popova, I. Y. Ho, N. T. Pauli, C. J. H. Dunand, W. M. Taylor, S. Lim, M. Huang, X. Qu, J.-H. Lee, M. Salgado-Ferrer, F. Krammer, P. Palese, J. Wrammert, R. Ahmed, P. C. Wilson, Immune history profoundly affects broadly protective B cell responses to influenza. *Science translational medicine*. **7**, 316ra192-316ra192 (2015).

59. G. Bajic, C. E. Poel, M. Kuraoka, A. G. Schmidt, M. C. Carroll, G. Kelsoe, S. C. Harrison, Autoreactivity profiles of influenza hemagglutinin broadly neutralizing antibodies. *Nature Publishing Group*. **9**, 1–9 (2019).

60. A. H. Ellebedy, F. Krammer, G.-M. Li, M. S. Miller, C. Chiu, J. Wrammert, C. Y. Chang, C. W. Davis, M. McCausland, R. Elbein, S. Edupuganti, P. Spearman, S. F. Andrews, P. C. Wilson, A. García-Sastre, M. J. Mulligan, A. K. Mehta, P. Palese, R. Ahmed, Induction of broadly cross-reactive antibody responses to the influenza HA stem region following H5N1 vaccination in humans. *Proc National Acad Sci*. **111**, 13133–13138 (2014).

61. L. Liu, R. Nachbagauer, L. Zhu, Y. Huang, X. Xie, S. Jin, A. Zhang, Y. Wan, A. Hirsh, D. Tian, X. Shi, Z. Dong, S. Yuan, Y. Hu, F. Krammer, X. Zhang, J. Xu, Induction of Broadly Cross-Reactive Stalk-Specific Antibody Responses to Influenza Group 1 and Group 2 Hemagglutinins by Natural H7N9 Virus Infection in Humans. *J Infect Dis*. **215**, 518–528 (2017).

62. S. F. Andrews, M. G. Joyce, M. J. Chambers, R. A. Gillespie, M. Kanekiyo, K. Leung, E. S.

Yang, Y. Tsybovsky, A. K. Wheatley, M. C. Crank, J. C. Boyington, M. S. Prabhakaran, S. R. Narpala, X. Chen, R. T. Bailer, G. Chen, E. Coates, P. D. Kwong, R. A. Koup, J. R. Mascola, B. S. Graham, J. E. Ledgerwood, A. B. McDermott, Preferential induction of cross-group influenza A hemagglutinin stem-specific memory B cells after H7N9 immunization in humans. *Sci Immunol.* **2**, eaan2676 (2017).

63. H. L. Dugan, J. J. Guthmiller, P. Arevalo, M. Huang, Y.-Q. Chen, K. E. Neu, C. Henry, N.-Y. Zheng, L. Y.-L. Lan, M. E. Tepora, O. Stovicek, D. Bitar, A.-K. E. Palm, C. T. Stamper, S. Changrob, H. A. Utset, L. Coughlan, F. Krammer, S. Cobey, P. C. Wilson, Preexisting immunity shapes distinct antibody landscapes after influenza virus infection and vaccination in humans. *Sci Transl Med.* **12**, eabd3601 (2020).

64. I. Skountzou, D. G. Koutsonanos, J. H. Kim, R. Powers, L. Satyabhama, F. Masseoud, W. C. Weldon, M. del P. Martin, R. S. Mittler, R. Compans, J. Jacob, Immunity to Pre-1950 H1N1 Influenza Viruses Confers Cross-Protection against the Pandemic Swine-Origin 2009 A (H1N1) Influenza Virus. *J Immunol.* **185**, 1642–1649 (2010).

65. J. M. Fonville, S. H. Wilks, S. L. James, A. Fox, M. Ventresca, M. Aban, L. Xue, T. C. Jones, L. N. M. H., P. Q. T., T. N. D., Y. Wong, A. Mosterin, L. C. Katzelnick, D. Labonte, L. T. T., G. van der Net, E. Skepner, C. A. Russell, T. D. Kaplan, G. F. Rimmelzwaan, N. Masurel, J. C. de Jong, A. Palache, W. E. P. Beyer, L. Q. M., N. T. H., H. F. L. Wertheim, A. C. Hurt, A. D. M. E. Osterhaus, I. G. Barr, R. A. M. Fouchier, P. W. Horby, D. J. Smith, Antibody landscapes after influenza virus infection or vaccination. *Science.* **346**, 996–1000 (2014).

66. K. M. Gostic, M. Ambrose, M. Worobey, J. O. Lloyd-Smith, Potent protection against H5N1 and H7N9 influenza via childhood hemagglutinin imprinting. *Science.* **354**, 1–6 (2016).

67. J. Wrammert, D. Koutsonanos, G.-M. Li, S. Edupuganti, J. Sui, M. Morrissey, M. McCausland, I. Skountzou, M. Hornig, W. I. Lipkin, A. Mehta, B. Razavi, C. D. Rio, N.-Y. Zheng, J.-H. Lee, M. Huang, Z. Ali, K. Kaur, S. Andrews, R. R. Amara, Y. Wang, S. R. Das, C. D. O'Donnell, J. W. Yewdell, K. Subbarao, W. A. Marasco, M. J. Mulligan, R. Compans, R. Ahmed, P. C. Wilson,

Broadly cross-reactive antibodies dominate the human B cell response against 2009 pandemic H1N1 influenza virus infection. *J Exp Med.* **208**, 181–193 (2011).

68. L. Mesin, A. Schiepers, J. Ersching, A. Barbulescu, C. B. Cavazzoni, A. Angelini, T. Okada, T. Kurosaki, G. D. Victora, Restricted Clonality and Limited Germinal Center Reentry Characterize Memory B Cell Reactivation by Boosting. *Cell.* **180**, 1–27 (2019).

69. J. L. Nayak, S. Alam, A. J. Sant, Cutting Edge: Heterosubtypic Influenza Infection Antagonizes Elicitation of Immunological Reactivity to Hemagglutinin. *J Immunol.* **191**, 1001–1005 (2013).

70. D. F. Hoft, K. Lottenbach, J. B. Goll, H. Hill, P. L. Winokur, S. M. Patel, R. C. Brady, W. H. Chen, K. Edwards, C. B. Creech, S. E. Frey, T. P. Blevins, R. Salomon, R. B. Belshe, Priming Vaccination With Influenza Virus H5 Hemagglutinin Antigen Significantly Increases the Duration of T cell Responses Induced by a Heterologous H5 Booster Vaccination. *J Infect Dis.* **214**, 1020–1029 (2016).

71. K. A. Richards, J. Nayak, F. A. Chaves, A. DiPiazza, Z. A. G. Knowlden, S. Alam, J. J. Treanor, A. J. Sant, Seasonal Influenza Can Poise Hosts for CD4 T-Cell Immunity to H7N9 Avian Influenza. *J Infect Dis.* **212**, 86–94 (2015).

72. H.-X. Tan, S. Jegaskanda, J. A. Juno, R. Esterbauer, J. Wong, H. G. Kelly, Y. Liu, D. Tilmanis, A. C. Hurt, J. W. Yewdell, S. J. Kent, A. K. Wheatley, Subdominance and poor intrinsic immunogenicity limit humoral immunity targeting influenza HA stem. *Journal of Clinical Investigation.* **129**, 850–862 (2019).

73. W.-C. Liu, R. Nachbagauer, D. Stadlbauer, A. Solórzano, F. Berlanda-Scorza, A. García-Sastre, P. Palese, F. Krammer, R. A. Albrecht, Sequential Immunization With Live-Attenuated Chimeric Hemagglutinin-Based Vaccines Confers Heterosubtypic Immunity Against Influenza A Viruses in a Preclinical Ferret Model. *Frontiers in Immunology.* **10**, 634–18 (2019).

74. M. Kanekiyo, M. G. Joyce, R. A. Gillespie, J. R. Gallagher, S. F. Andrews, H. M. Yassine, A.

K. Wheatley, B. E. Fisher, D. R. Ambrozak, A. Creanga, K. Leung, E. S. Yang, S. Boyoglu-Barnum, I. S. Georgiev, Y. Tsybovsky, M. S. Prabhakaran, H. Andersen, W.-P. Kong, U. Baxa, K. L. Zephir, J. E. Ledgerwood, R. A. Koup, P. D. Kwong, A. K. Harris, A. B. McDermott, J. R. Mascola, B. S. Graham, Mosaic nanoparticle display of diverse influenza virus hemagglutinins elicits broad B cell responses. *Nature Immunology*. 20, 1–18 (2019).

75. S. Boyoglu-Barnum, D. Ellis, R. A. Gillespie, G. B. Hutchinson, Y.-J. Park, S. M. Moin, O. J. Acton, R. Ravichandran, M. Murphy, D. Pettie, N. Matheson, L. Carter, A. Creanga, M. J. Watson, S. Kephart, S. Ataca, J. R. Valle, G. Ueda, M. C. Crank, L. Stewart, K. K. Lee, M. Guttman, D. Baker, J. R. Mascola, D. Veesler, B. S. Graham, N. P. King, M. Kanekiyo, Quadrivalent influenza nanoparticle vaccines induce broad protection. *Nature*. **592**, 623–628 (2021).

76. W. Kamitani, K. Narayanan, C. Huang, K. Lokugamage, T. Ikegami, N. Ito, H. Kubo, S. Makino, Severe acute respiratory syndrome coronavirus nsp1 protein suppresses host gene expression by promoting host mRNA degradation. *Proc National Acad Sci*. 103, 12885–12890 (2006).

77. J. D. Cherry, P. Krogstad, SARS: The First Pandemic of the 21st Century. *Pediatr Res*. **56**, 1–5 (2004).

78. E. de Wit, N. van Doremalen, D. Falzarano, V. J. Munster, SARS and MERS: recent insights into emerging coronaviruses. *Nat Rev Microbiol*. **14**, 523–534 (2016).

79. W. Li, Z. Shi, M. Yu, W. Ren, C. Smith, J. H. Epstein, H. Wang, G. Crameri, Z. Hu, H. Zhang, J. Zhang, J. McEachern, H. Field, P. Daszak, B. T. Eaton, S. Zhang, L.-F. Wang, Bats Are Natural Reservoirs of SARS-Like Coronaviruses. *Science*. **310**, 676–679 (2005).

80. X.-Y. Ge, J.-L. Li, X.-L. Yang, A. A. Chmura, G. Zhu, J. H. Epstein, J. K. Mazet, B. Hu, W. Zhang, C. Peng, Y.-J. Zhang, C.-M. Luo, B. Tan, N. Wang, Y. Zhu, G. Crameri, S.-Y. Zhang, L.-F. Wang, P. Daszak, Z.-L. Shi, Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. *Nature*. 503, 535–538 (2013).

81. V. D. Menachery, B. L. Yount, K. Debbink, S. Agnihothram, L. E. Gralinski, J. A. Plante, R. L.

Graham, T. Scobey, X.-Y. Ge, E. F. Donaldson, S. H. Randell, A. Lanzavecchia, W. A. Marasco, Z.-L. Shi, R. S. Baric, A SARS-like cluster of circulating bat coronaviruses shows potential for human emergence. *Nat Med.* **21**, 1508–1513 (2015).

82. V. D. Menachery, B. L. Yount, A. C. Sims, K. Debbink, S. S. Agnihothram, L. E. Gralinski, R. L. Graham, T. Scobey, J. A. Plante, S. R. Royal, J. Swanstrom, T. P. Sheahan, R. J. Pickles, D. Corti, S. H. Randell, A. Lanzavecchia, W. A. Marasco, R. S. Baric, SARS-like WIV1-CoV poised for human emergence. *Proc National Acad Sci.* **113**, 3048–3053 (2016).

83. N. Wang, S.-Y. Li, X.-L. Yang, H.-M. Huang, Y.-J. Zhang, H. Guo, C.-M. Luo, M. Miller, G. Zhu, A. A. Chmura, E. Hagan, J.-H. Zhou, Y.-Z. Zhang, L.-F. Wang, P. Daszak, Z.-L. Shi, Serological Evidence of Bat SARS-Related Coronavirus Infection in Humans, China. *Virologica Sinica.* **33**, 104–107 (2018).

84. T. Carvalho, F. Krammer, A. Iwasaki, The first 12 months of COVID-19: a timeline of immunological insights. *Nat Rev Immunol.* **21**, 245–256 (2021).

85. Q. Li, X. Guan, P. Wu, X. Wang, L. Zhou, Y. Tong, R. Ren, K. S. M. Leung, E. H. Y. Lau, J. Y. Wong, X. Xing, N. Xiang, Y. Wu, C. Li, Q. Chen, D. Li, T. Liu, J. Zhao, M. Liu, W. Tu, C. Chen, L. Jin, R. Yang, Q. Wang, S. Zhou, R. Wang, H. Liu, Y. Luo, Y. Liu, G. Shao, H. Li, Z. Tao, Y. Yang, Z. Deng, B. Liu, Z. Ma, Y. Zhang, G. Shi, T. T. Y. Lam, J. T. Wu, G. F. Gao, B. J. Cowling, B. Yang, G. M. Leung, Z. Feng, Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. *New Engl J Med.* **382**, 1199–1207 (2020).

86. Z. Yuan, Y. Xiao, Z. Dai, J. Huang, Z. Zhang, Y. Chen, Modelling the effects of Wuhan's lockdown during COVID-19, China. *B World Health Organ.* **98**, 484–494 (n.d.).

87. M. F. Boni, P. Lemey, X. Jiang, T. T.-Y. Lam, B. W. Perry, T. A. Castoe, A. Rambaut, D. L. Robertson, Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic. *Nat Microbiol.* **5**, 1408–1417 (2020).

88. E. Dong, H. Du, L. Gardner, An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis.* **20**, 533–534 (2020).

89. S. Sanche, Y. T. Lin, C. Xu, E. Romero-Severson, N. Hengartner, R. Ke, Early Release - High Contagiousness and Rapid Spread of Severe Acute Respiratory Syndrome Coronavirus 2 - Volume 26, Number 7—July 2020 - Emerging Infectious Diseases journal - CDC. *Emerg Infect Dis.* 26, 1470–1477 (n.d.).

90. X. He, E. H. Y. Lau, P. Wu, X. Deng, J. Wang, X. Hao, Y. C. Lau, J. Y. Wong, Y. Guan, X. Tan, X. Mo, Y. Chen, B. Liao, W. Chen, F. Hu, Q. Zhang, M. Zhong, Y. Wu, L. Zhao, F. Zhang, B. J. Cowling, F. Li, G. M. Leung, Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nat Med.* 26, 672–675 (2020).

91. N. H. L. Leung, Transmissibility and transmission of respiratory viruses. *Nat Rev Microbiol.* 1–18 (2021).

92. N. Askitas, K. Tatsiramos, B. Verheyden, Estimating worldwide effects of non-pharmaceutical interventions on COVID-19 incidence and population mobility patterns using a multiple-event study. *Sci Rep-uk.* 11, 1972 (2021).

93. W. J. Wiersinga, A. Rhodes, A. C. Cheng, S. J. Peacock, H. C. Prescott, Pathophysiology, Transmission, Diagnosis, and Treatment of Coronavirus Disease 2019 (COVID-19). *Jama.* **324**, 782–793 (2020).

94. A. P. West, J. O. Wertheim, J. C. Wang, T. I. Vasylyeva, J. L. Havens, M. A. Chowdhury, E. Gonzalez, C. E. Fang, S. S. D. Lonardo, S. Hughes, J. L. Rakeman, H. H. Lee, C. O. Barnes, P. N. P. Gnanapragasam, Z. Yang, C. Gaebler, M. Caskey, M. C. Nussenzweig, J. R. Keeffe, P. J. Bjorkman, *Biorxiv*, in press, doi:10.1101/2021.02.14.431043.

95. A. J. Greaney, T. N. Starr, P. Gilchuk, S. J. Zost, E. Binshtein, A. N. Loes, S. K. Hilton, J. Huddleston, R. Eguia, K. H. D. Crawford, A. S. Dingens, R. S. Nargi, R. E. Sutton, N. Suryadevara, P. W. Rothlauf, Z. Liu, S. P. J. Whelan, R. H. Carnahan, J. E. Crowe, J. D. Bloom, Complete Mapping of Mutations to the SARS-CoV-2 Spike Receptor-Binding Domain that Escape Antibody Recognition. *Cell Host Microbe.* 29, 44-57.e9 (2021).

96. M. S. Graham, C. H. Sudre, A. May, M. Antonelli, B. Murray, T. Varsavsky, K. Kläser, L. S.

Canas, E. Molteni, M. Modat, D. A. Drew, L. H. Nguyen, L. Polidori, S. Selvachandran, C. Hu, J. Capdevila, C.-19 G. U. (COG-U. Consortium, C. Koshy, A. Ash, E. Wise, N. Moore, M. Mori, N. Cortes, J. Lynch, S. Kidd, D. J. Fairley, T. Curran, J. P. McKenna, H. Adams, C. Fraser, T. Golubchik, D. Bonsall, M. O. Hassan-Ibrahim, C. S. Malone, B. J. Cogger, M. Wantoch, N. Reynolds, B. Warne, J. Maksimovic, K. Spellman, K. McCluggage, M. John, R. Beer, S. Afifi, S. Morgan, A. Marchbank, A. Price, C. Kitchen, H. Gulliver, I. Merrick, J. Southgate, M. Guest, R. Munn, T. Workman, T. R. Connor, W. Fuller, C. Bresner, L. B. Snell, A. Patel, T. Charalampous, G. Nebbia, R. Batra, J. Edgeworth, S. C. Robson, A. H. Beckett, D. M. Aanensen, A. P. Underwood, C. A. Yeats, K. Abudahab, B. E. Taylor, M. Menegazzo, G. Clark, W. Smith, M. Khakh, V. M. Fleming, M. M. Lister, H. C. Howson-Wells, L. Berry, T. Boswell, A. Joseph, I. Willingham, C. Jones, C. Holmes, P. Bird, T. Helmer, K. Fallon, J. Tang, V. Raviprakash, S. Campbell, N. Sheriff, V. Blakey, L.-A. Williams, M. W. Loose, N. Holmes, C. Moore, M. Carlile, V. Wright, F. Sang, J. Debebe, F. Coll, A. W. Signell, G. Betancor, H. D. Wilson, S. Eldirdiri, A. Kenyon, T. Davis, O. G. Pybus, L. du Plessis, A. E. Zarebski, J. Raghwani, M. U. Kraemer, S. Francois, S. W. Attwood, T. I. Vasylyeva, M. E. Zamudio, B. Gutierrez, M. E. Torok, W. L. Hamilton, I. G. Goodfellow, G. Hall, A. S. Jahun, Y. Chaudhry, M. Hosmillo, M. L. Pinckert, I. Georgana, S. Moses, H. Lowe, L. Bedford, J. Moore, S. Stonehouse, C. L. Fisher, A. R. Awan, J. BoYes, J. Breuer, K. A. Harris, J. R. Brown, D. Shah, L. Atkinson, J. C. Lee, N. Storey, F. Flaviani, A. Alcolea-Medina, R. Williams, G. Vernet, M. R. Chapman, L. J. Levett, J. Heaney, W. Chatterton, M. Pusok, L. Xu-McCrae, D. L. Smith, M. Bashton, G. R. Young, A. Holmes, P. A. Randell, A. Cox, P. Madona, F. Bolt, J. Price, S. Mookerjee, M. Ragonnet-Cronin, F. F. Nascimento, D. Jorgensen, I. Siveroni, R. Johnson, O. Boyd, L. Geidelberg, E. M. Volz, A. Rowan, G. P. Taylor, K. L. Smollett, N. J. Loman, J. Quick, C. McMurray, J. Stockton, S. Nicholls, W. Rowe, R. Poplawski, A. McNally, R. T. M. Nunez, J. Mason, T. I. Robinson, E. O'Toole, J. Watts, C. Breen, A. Cowell, G. Sluga, N. W. Machin, S. S. Y. Ahmad, R. P. George, F. Halstead, V. Sivaprakasam, W. Hogsden, C. J. Illingworth, C. Jackson, E. C. Thompson, J. G. Shepherd, P. Asamaphan, M. O. Niebel, K. K. Li, R. N. Shah, N. G. Jesudason, L.

Tong, A. Broos, D. Mair, J. Nichols, S. N. Carmichael, K. Nomikou, E. Aranday-Cortes, N. Johnson, I. Starinskij, A. da S. Filipe, D. L. Robertson, R. J. Orton, J. Hughes, S. Vattipally, J. B. Singer, S. Nickbakhsh, A. D. Hale, L. R. Macfarlane-Smith, K. L. Harper, H. Carden, Y. Taha, B. A. Payne, S. Burton-Fanning, S. Waugh, J. Collins, G. Eltringham, S. Rushton, S. O'Brien, A. Bradley, A. Maclean, G. Mollett, R. Blacow, K. E. Templeton, M. P. McHugh, R. Dewar, E. Wastenge, S. Dervisevic, R. Stanley, E. J. Meader, L. Coupland, L. Smith, C. Graham, E. Barton, D. Padgett, G. Scott, E. Swindells, J. Greenaway, A. Nelson, C. M. McCann, W. C. Yew, M. Andersson, T. Peto, A. Justice, D. Eyre, D. Crook, T. J. Sloan, N. Duckworth, S. Walsh, A. J. Chauhan, S. Glaysher, K. Bicknell, S. Wyllie, S. Elliott, A. Lloyd, R. Impey, N. Levene, L. Monaghan, D. T. Bradley, T. Wyatt, E. Allara, C. Pearson, H. Osman, A. Bosworth, E. Robinson, P. Muir, I. B. Vipond, R. Hopes, H. M. Pymont, S. Hutchings, M. D. Curran, S. Parmar, A. Lackenby, T. Mbisa, S. Platt, S. Miah, D. Bibby, C. Manso, J. Hubb, M. Chand, G. Dabrera, M. Ramsay, D. Bradshaw, A. Thornton, R. Myers, U. Schaefer, N. Groves, E. Gallagher, D. Lee, D. Williams, N. Ellaby, I. Harrison, H. Hartman, N. Manesis, V. Patel, C. Bishop, V. Chalker, J. Ledesma, K. A. Twohig, M. T. G. Holden, S. Shaaban, A. Birchley, A. Adams, A. Davies, A. Gaskin, A. Plimmer, B. Gatica-Wilcox, C. McKerr, C. Moore, C. Williams, D. Heyburn, E. D. Lacy, E. Hilvers, F. Downing, G. Shankar, H. Jones, H. Asad, J. Coombes, J. Watkins, J. M. Evans, L. Fina, L. Gifford, L. Gilbert, L. Graham, M. Perry, M. Morgan, M. Bull, M. Cronin, N. Pacchiarini, N. Craine, R. Jones, R. Howe, S. Corden, S. Rey, S. Kumziene-SummerhaYes, S. Taylor, S. Cottrell, S. Jones, S. Edwards, J. O'Grady, A. J. Page, A. E. Mather, D. J. Baker, S. Rudder, A. Aydin, G. L. Kay, A. J. Trotter, N.-F. Alikhan, L. de O. Martins, T. Le-Viet, L. Meadows, A. Casey, L. Ratcliffe, D. A. Simpson, Z. Molnar, T. Thompson, E. Acheson, J. A. Masoli, B. A. Knight, S. Ellard, C. Auckland, C. R. Jones, T. W. Mahungu, D. Irish-Tavares, T. Haque, J. Hart, E. Witele, M. L. Fenton, A. Dadrah, A. Symmonds, T. Saluja, Y. Bourgeois, G. P. Scarlett, K. F. Loveson, S. Goudarzi, C. Fearn, K. Cook, H. Dent, H. Paul, D. G. Partridge, M. Raza, C. Evans, K. Johnson, S. Liggett, P. Baker, S. Bonner, S. Essex, R. A. Lyons, K. Saeed, A. I. K. Mahanama, B. Samaraweera, S. Silveira, E.

Pelosi, E. Wilson-Davies, R. J. Williams, M. Kristiansen, S. Roy, C. A. Williams, M. Cotic, N. Bayzid, A. P. Westhorpe, J. A. Hartley, R. Jannoo, H. L. Lowe, A. Karamani, L. Ensell, J. A. Prieto, S. Jeremiah, D. Grammatopoulos, S. Pandey, L. Berry, K. Jones, A. Richter, A. Beggs, A. Best, B. Percival, J. Mirza, O. Megram, M. Mayhew, L. Crawford, F. Ashcroft, E. Moles-Garcia, N. Cumley, C. P. Smith, G. Bucca, A. R. Hesketh, B. Blane, S. T. Girgis, D. Leek, S. Sridhar, S. Forrest, C. Cormie, H. K. Gill, J. Dias, E. E. Higginson, M. Maes, J. Young, L. M. Kermack, R. K. Gupta, C. Ludden, S. J. Peacock, S. Palmer, C. M. Churcher, N. F. Hadjirin, A. M. Carabelli, E. Brooks, K. S. Smith, K. Galai, G. M. McManus, C. Ruis, R. K. Davidson, A. Rambaut, T. Williams, C. E. Balcazar, M. D. Gallagher, Á. O'Toole, S. Rooke, V. Hill, K. A. Williamson, T. D. Stanton, S. L. Michell, C. M. Bewshea, B. Temperton, M. L. Michelsen, J. Warwick-Dugdale, R. Manley, A. Farbos, J. W. Harrison, C. M. Sambles, D. J. Studholme, A. R. Jeffries, A. C. Darby, J. A. Hiscox, S. Paterson, M. Iturriza-Gomara, K. A. Jackson, A. O. Lucaci, E. E. Vamos, M. Hughes, L. Rainbow, R. Eccles, C. Nelson, M. Whitehead, L. Turtle, S. T. Haldenby, R. Gregory, M. Gemmell, C. Wierzbicki, H. J. Webster, T. I. de Silva, N. Smith, A. Angyal, B. B. Lindsey, D. C. Groves, L. R. Green, D. Wang, T. M. Freeman, M. D. Parker, A. J. Keeley, P. J. Parsons, R. M. Tucker, R. Brown, M. Wyles, M. Whiteley, P. Zhang, M. Gallis, S. F. Louka, C. Constantiniou, M. Unnikrishnan, S. Ott, J. K. J. Cheng, H. E. Bridgewater, L. R. Frost, G. Taylor-Joyce, R. Stark, L. Baxter, M. T. Alam, P. E. Brown, D. Aggarwal, A. C. Cerdá, T. V. Merrill, R. E. Wilson, P. C. McClure, J. G. Chappell, T. Tsoleridis, J. Ball, D. Buck, J. A. Todd, A. Green, A. Trebes, G. MacIntyre-Cockett, M. de Cesare, A. Alderton, R. Amato, C. V. Ariani, M. A. Beale, C. Beaver, K. L. Bellis, E. Betteridge, J. Bonfield, J. Danesh, M. J. Dorman, E. Drury, B. W. Farr, L. Foulser, S. Goncalves, S. Goodwin, M. Gourtovaia, E. M. Harrison, D. K. Jackson, D. Jamrozy, I. Johnston, L. Kane, S. Kay, J.-P. Keatley, D. Kwiatkowski, C. F. Langford, M. Lawniczak, L. Letchford, R. Livett, S. Lo, I. Martincorena, S. McGuigan, R. Nelson, S. Palmer, N. R. Park, M. Patel, L. Prestwood, C. Puethé, M. A. Quail, S. Rajatileka, C. Scott, L. Shirley, J. Sillitoe, M. H. S. Chapman, S. A. Thurston, G. Tonkin-Hill, D. Weldon, D. Rajan, I. F. Bronner, L. Aigrain, N. M. Redshaw, S. V. Lensing, R. Davies,

A. Whitwham, J. Liddle, K. Lewis, J. M. Tovar-Corona, S. Leonard, J. Durham, A. R. Bassett, S. McCarthy, R. J. Moll, K. James, K. Oliver, A. Makunin, J. Barrett, R. N. Gunson, A. Hammers, A. T. Chan, J. Wolf, T. D. Spector, C. J. Steves, S. Ourselin, Changes in symptomatology, reinfection, and transmissibility associated with the SARS-CoV-2 variant B.1.1.7: an ecological study. *Lancet Public Heal.* **6**, e335–e345 (2021).

97. P. V'kovski, A. Kratzel, S. Steiner, H. Stalder, V. Thiel, Coronavirus biology and replication: implications for SARS-CoV-2. *Nat Rev Microbiol.* **19**, 155–170 (2021).

98. D. Kim, J.-Y. Lee, J.-S. Yang, J. W. Kim, V. N. Kim, H. Chang, The Architecture of SARS-CoV-2 Transcriptome. *Cell.* **181**, 914-921.e10 (2020).

99. K. Schubert, E. D. Karousis, A. Jomaa, A. Scaiola, B. Echeverria, L.-A. Gurzeler, M. Leibundgut, V. Thiel, O. Mühlmann, N. Ban, SARS-CoV-2 Nsp1 binds the ribosomal mRNA channel to inhibit translation. *Nat Struct Mol Biol.* **27**, 959–966 (2020).

100. D. M. Skowronski, C. Astell, R. C. Brunham, D. E. Low, M. Petric, R. L. Roper, P. J. Talbot, T. Tam, L. Babiuk, Severe Acute Respiratory Syndrome (SARS): A Year in Review. *Medicine.* **56**, 357–381 (2005).

101. Z. Jin, X. Du, Y. Xu, Y. Deng, M. Liu, Y. Zhao, B. Zhang, X. Li, L. Zhang, C. Peng, Y. Duan, J. Yu, L. Wang, K. Yang, F. Liu, R. Jiang, X. Yang, T. You, X. Liu, X. Yang, F. Bai, H. Liu, X. Liu, L. W. Guddat, W. Xu, G. Xiao, C. Qin, Z. Shi, H. Jiang, Z. Rao, H. Yang, Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. *Nature.* **582**, 289–293 (2020).

102. J. Qiao, Y.-S. Li, R. Zeng, F.-L. Liu, R.-H. Luo, C. Huang, Y.-F. Wang, J. Zhang, B. Quan, C. Shen, X. Mao, X. Liu, W. Sun, W. Yang, X. Ni, K. Wang, L. Xu, Z.-L. Duan, Q.-C. Zou, H.-L. Zhang, W. Qu, Y.-H.-P. Long, M.-H. Li, R.-C. Yang, X. Liu, J. You, Y. Zhou, R. Yao, W.-P. Li, J.-M. Liu, P. Chen, Y. Liu, G.-F. Lin, X. Yang, J. Zou, L. Li, Y. Hu, G.-W. Lu, W.-M. Li, Y.-Q. Wei, Y.-T. Zheng, J. Lei, S. Yang, SARS-CoV-2 Mpro inhibitors with antiviral activity in a transgenic mouse model. *Science.* **371**, 1374–1378 (2021).

103. A. C. Walls, Y.-J. Park, M. A. Tortorici, A. Wall, A. T. McGuire, D. Veesler, Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell*, 1–19 (2020).

104. M. Hoffmann, H. Kleine-Weber, S. Schroeder, N. Krüger, T. Herrler, S. Erichsen, T. S. Schiergens, G. Herrler, N.-H. Wu, A. Nitsche, M. A. Müller, C. Drosten, S. Pöhlmann, SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell*, 1–19 (2020).

105. M. Letko, A. Marzi, V. Munster, Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. *Nature Microbiology*. **11**, 1–17 (2020).

106. J. Lan, J. Ge, J. Yu, S. Shan, H. Zhou, S. Fan, Q. Zhang, X. Shi, Q. Wang, L. Zhang, X. Wang, Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature*. **581**, 215–220 (2020).

107. R. Yan, Y. Zhang, Y. Li, L. Xia, Y. Guo, Q. Zhou, Structural basis for the recognition of the SARS-CoV-2 by full-length human ACE2. *Science*, eabb2762-10 (2020).

108. S. Liu, P. Selvaraj, C. Z. Lien, I. A. Nunez, W. W. Wu, C.-K. Chou, T. T. Wang, The PRRA insert at the S1/S2 site modulates cellular tropism of SARS-CoV-2 and ACE2 usage by the closely related Bat RaTG13. *J Virol* (2021), doi:10.1128/jvi.01751-20.

109. J. Damas, G. M. Hughes, K. C. Keough, C. A. Painter, N. S. Persky, M. Corbo, M. Hiller, K.-P. Koepfli, A. R. Pfenning, H. Zhao, D. P. Genereux, R. Swofford, K. S. Pollard, O. A. Ryder, M. T. Nweeia, K. Lindblad-Toh, E. C. Teeling, E. K. Karlsson, H. A. Lewin, Broad host range of SARS-CoV-2 predicted by comparative and structural analysis of ACE2 in vertebrates. *Proc National Acad Sci*. **117**, 22311–22322 (2020).

110. B. J. Bosch, W. Bartelink, P. J. M. Rottier, Cathepsin L Functionally Cleaves the Severe Acute Respiratory Syndrome Coronavirus Class I Fusion Protein Upstream of Rather than Adjacent to the Fusion Peptide . *J Virol*. **82**, 8887–8890 (2008).

111. S. Matsuyama, N. Nagata, K. Shirato, M. Kawase, M. Takeda, F. Taguchi, Efficient Activation of the Severe Acute Respiratory Syndrome Coronavirus Spike Protein by the Transmembrane Protease TMPRSS2 *v*. *J Virol.* **84**, 12658–12664 (2010).

112. S. Belouzard, V. C. Chu, G. R. Whittaker, Activation of the SARS coronavirus spike protein via sequential proteolytic cleavage at two distinct sites. *Proc National Acad Sci.* **106**, 5871–5876 (2009).

113. N. Iwata-Yoshikawa, T. Okamura, Y. Shimizu, H. Hasegawa, M. Takeda, N. Nagata, TMPRSS2 Contributes to Virus Spread and Immunopathology in the Airways of Murine Models after Coronavirus Infection. *Journal of Virology.* **93**, 1953–15 (2019).

114. M. Hoffmann, H. Kleine-Weber, S. Schroeder, N. Krüger, T. Herrler, S. Erichsen, T. S. Schiergens, G. Herrler, N.-H. Wu, A. Nitsche, M. A. Müller, C. Drosten, S. Pöhlmann, SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell.* **181**, 271-280.e8 (2020).

115. T. Ou, H. Mou, L. Zhang, A. Ojha, H. Choe, M. Farzan, Hydroxychloroquine-mediated inhibition of SARS-CoV-2 entry is attenuated by TMPRSS2. *Plos Pathog.* **17**, e1009212 (2021).

116. M. Hoffmann, S. Schroeder, H. Kleine-Weber, M. A. Müller, C. Drosten, S. Pöhlmann, Nafamostat Mesylate Blocks Activation of SARS-CoV-2: New Treatment Option for COVID-19. *Antimicrob Agents Ch.* **64** (2020), doi:10.1128/aac.00754-20.

117. W.-J. Shieh, C.-H. Hsiao, C. D. Paddock, J. Guarner, C. S. Goldsmith, K. Tatti, M. Packard, L. Mueller, M.-Z. Wu, P. Rollin, I.-J. Su, S. R. Zaki, Immunohistochemical, *in situ* hybridization, and ultrastructural localization of SARS-associated coronavirus in lung of a fatal case of severe acute respiratory syndrome in Taiwan. *Hum Pathol.* **36**, 303–309 (2005).

118. I. Hamming, W. Timens, M. Bulthuis, A. Lely, G. Navis, H. van Goor, Tissue distribution of ACE2 protein, the functional receptor for SARS coronavirus. A first step in understanding SARS pathogenesis. *J Pathology.* **203**, 631–637 (2004).

119. R. Wölfel, V. M. Corman, W. Guggemos, M. Seilmaier, S. Zange, M. A. Müller, D. Niemeyer, T. C. Jones, P. Vollmar, C. Rothe, M. Hoelscher, T. Bleicker, S. Brünink, J. Schneider, R. Ehmann, K. Zwirglmaier, C. Drosten, C. Wendtner, Virological assessment of hospitalized patients with COVID-2019. *Nature*. **581**, 465–469 (2020).

120. S. Ozono, Y. Zhang, H. Ode, K. Sano, T. S. Tan, K. Imai, K. Miyoshi, S. Kishigami, T. Ueno, Y. Iwatani, T. Suzuki, K. Tokunaga, SARS-CoV-2 D614G spike mutation increases entry efficiency with enhanced ACE2-binding affinity. *Nat Commun*. **12**, 848 (2021).

121. J. L. Daly, B. Simonetti, K. Klein, K.-E. Chen, M. K. Williamson, C. Antón-Plágaro, D. K. Shoemark, L. Simón-Gracia, M. Bauer, R. Hollandi, U. F. Greber, P. Horvath, R. B. Sessions, A. Helenius, J. A. Hiscox, T. Teesalu, D. A. Matthews, A. D. Davidson, B. M. Collins, P. J. Cullen, Y. Yamauchi, Neuropilin-1 is a host factor for SARS-CoV-2 infection. *Science*. **370**, 861–865 (2020).

122. T. P. Peacock, D. H. Goldhill, J. Zhou, L. Baillon, R. Frise, O. C. Swann, R. Kugathasan, R. Penn, J. C. Brown, R. Y. Sanchez-David, L. Braga, M. K. Williamson, J. A. Hassard, E. Staller, B. Hanley, M. Osborn, M. Giacca, A. D. Davidson, D. A. Matthews, W. S. Barclay, The furin cleavage site in the SARS-CoV-2 spike protein is required for transmission in ferrets. *Nat Microbiol*, 1–11 (2021).

123. B. A. Johnson, X. Xie, A. L. Bailey, B. Kalveram, K. G. Lokugamage, A. Muruato, J. Zou, X. Zhang, T. Juelich, J. K. Smith, L. Zhang, N. Bopp, C. Schindewolf, M. Vu, A. Vanderheiden, E. S. Winkler, D. Swetnam, J. A. Plante, P. Aguilar, K. S. Plante, V. Popov, B. Lee, S. C. Weaver, M. S. Suthar, A. L. Routh, P. Ren, Z. Ku, Z. An, K. Debbink, M. S. Diamond, P.-Y. Shi, A. N. Freiberg, V. D. Menachery, Loss of furin cleavage site attenuates SARS-CoV-2 pathogenesis. *Nature*. **591**, 293–299 (2021).

124. J. K. Millet, G. R. Whittaker, Host cell entry of Middle East respiratory syndrome coronavirus after two-step, furin-mediated activation of the spike protein. *Proc National Acad Sci*. **111**, 15214–15219 (2014).

125. H. Zhou, X. Chen, T. Hu, J. Li, H. Song, Y. Liu, P. Wang, D. Liu, J. Yang, E. C. Holmes, A.

C. Hughes, Y. Bi, W. Shi, A novel bat coronavirus closely related to SARS-CoV-2 contains natural insertions at the S1/S2 cleavage site of the spike protein. *Curr Biol* (2020), doi:10.1016/j.cub.2020.05.023.

126. A. Grifoni, D. Weiskopf, S. I. Ramirez, J. Mateus, J. M. Dan, C. R. Moderbacher, S. A. Rawlings, A. Sutherland, L. Premkumar, R. S. Jadi, D. Marrama, A. M. de Silva, A. Frazier, A. F. Carrilin, J. A. Greenbaum, B. Peters, F. Krammer, D. M. Smith, S. Crotty, A. Sette, Targets of T Cell Responses to SARS-CoV-2 Coronavirus in Humans with COVID-19 Disease and Unexposed Individuals. *Cell*. 181, 1489-1501.e15 (2020).

127. J. Zuo, A. C. Dowell, H. Pearce, K. Verma, H. M. Long, J. Begum, F. Aiano, Z. Amin-Chowdhury, B. Hallis, L. Stapley, R. Borrow, E. Linley, S. Ahmad, B. Parker, A. Horsley, G. Amirthalingam, K. Brown, M. E. Ramsay, S. Ladhami, P. Moss, Robust SARS-CoV-2-specific T cell immunity is maintained at 6 months following primary infection. *Nat Immunol*, 1–7 (2021).

128. J. Mateus, A. Grifoni, A. Tarke, J. Sidney, S. I. Ramirez, J. M. Dan, Z. C. Burger, S. A. Rawlings, D. M. Smith, E. Phillips, S. Mallal, M. Lammers, P. Rubiro, L. Quiambao, A. Sutherland, E. D. Yu, R. da S. Antunes, J. Greenbaum, A. Frazier, A. J. Markmann, L. Premkumar, A. de Silva, B. Peters, S. Crotty, A. Sette, D. Weiskopf, Selective and cross-reactive SARS-CoV-2 T cell epitopes in unexposed humans. *Science*. 370, 89–94 (2020).

129. A. T. DiPiazza, B. S. Graham, T. J. Ruckwardt, T cell immunity to SARS-CoV-2 following natural infection and vaccination. *Biochem Biophys Res Co.* **538**, 211–217 (2020).

130. D. Mathew, J. R. Giles, A. E. Baxter, D. A. Oldridge, A. R. Greenplate, J. E. Wu, C. Alanio, L. Kuri-Cervantes, M. B. Pampena, K. D'Andrea, S. Manne, Z. Chen, Y. J. Huang, J. P. Reilly, A. R. Weisman, C. A. G. Ittner, O. Kuthuru, J. Dougherty, K. Nzingha, N. Han, J. Kim, A. Pattekar, E. C. Goodwin, E. M. Anderson, M. E. Weirick, S. Gouma, C. P. Arevalo, M. J. Bolton, F. Chen, S. F. Lacey, H. Ramage, S. Cherry, S. E. Hensley, S. A. Apostolidis, A. C. Huang, L. A. Vella, T. Up. C. P. Unit†, M. R. Betts, N. J. Meyer, E. J. Wherry, Deep immune profiling of COVID-19 patients reveals distinct immunotypes with therapeutic implications. *Science*, eabc8511 (2020).

131. R. R. Goel, S. A. Apostolidis, M. M. Painter, D. Mathew, A. Pattekar, O. Kuthuru, S. Gouma, P. Hicks, W. Meng, A. M. Rosenfeld, S. Dysinger, K. A. Lundgreen, L. Kuri-Cervantes, S. Adamski, A. Hicks, S. Korte, D. A. Oldridge, A. E. Baxter, J. R. Giles, M. E. Weirick, C. M. McAllister, J. Dougherty, S. Long, K. D'Andrea, J. T. Hamilton, M. R. Betts, E. T. L. Prak, P. Bates, S. E. Hensley, A. R. Greenplate, E. J. Wherry, Distinct antibody and memory B cell responses in SARS-CoV-2 naïve and recovered individuals following mRNA vaccination. *Sci Immunol.* **6**, eabi6950 (2021).

132. K. Röltgen, A. E. Powell, O. F. Wirz, B. A. Stevens, C. A. Hogan, J. Najeeb, M. Hunter, H. Wang, M. K. Sahoo, C. Huang, F. Yamamoto, M. Manohar, J. Manalac, A. R. Otrelo-Cardoso, T. D. Pham, A. Rustagi, A. J. Rogers, N. H. Shah, C. A. Blish, J. R. Cochran, T. S. Jardetzky, J. L. Zehnder, T. T. Wang, B. Narasimhan, S. Gombar, R. Tibshirani, K. C. Nadeau, P. S. Kim, B. A. Pinsky, S. D. Boyd, Defining the features and duration of antibody responses to SARS-CoV-2 infection associated with disease severity and outcome. *Sci Immunol.* **5**, eabe0240 (2020).

133. D. F. Robbiani, C. Gaebler, F. Muecksch, J. C. C. Lorenzi, Z. Wang, A. Cho, M. Agudelo, C. O. Barnes, A. Gazumyan, S. Finkin, T. Hägglöf, T. Y. Oliveira, C. Viant, A. Hurley, H.-H. Hoffmann, K. G. Millard, R. G. Kost, M. Cipolla, K. Gordon, F. Bianchini, S. T. Chen, V. Ramos, R. Patel, J. Dizon, I. Shimeliovich, P. Mendoza, H. Hartweger, L. Nogueira, M. Pack, J. Horowitz, F. Schmidt, Y. Weisblum, E. Michailidis, A. W. Ashbrook, E. Waltari, J. E. Pak, K. E. Huey-Tubman, N. Koranda, P. R. Hoffman, A. P. West, C. M. Rice, T. Hatziloannou, P. J. Bjorkman, P. D. Bieniasz, M. Caskey, M. C. Nussenzweig, Convergent antibody responses to SARS-CoV-2 in convalescent individuals. *Nature.* **584**, 437–442 (2020).

134. C. O. Barnes, A. P. West, K. E. Huey-Tubman, M. A. G. Hoffmann, N. G. Sharaf, P. R. Hoffman, N. Koranda, H. B. Gristick, C. Gaebler, F. Muecksch, J. C. C. Lorenzi, S. Finkin, T. Hägglöf, A. Hurley, K. G. Millard, Y. Weisblum, F. Schmidt, T. Hatziloannou, P. D. Bieniasz, M. Caskey, D. F. Robbiani, M. C. Nussenzweig, P. J. Bjorkman, Structures of Human Antibodies Bound to SARS-CoV-2 Spike Reveal Common Epitopes and Recurrent Features of Antibodies. *Cell.* **182**,

828-842.e16 (2020).

135. C. O. Barnes, C. A. Jette, M. E. Abernathy, K.-M. A. Dam, S. R. Esswein, H. B. Gristick, A. G. Malyutin, N. G. Sharaf, K. E. Huey-Tubman, Y. E. Lee, D. F. Robbiani, M. C. Nussenzweig, A. P. West, P. J. Bjorkman, SARS-CoV-2 neutralizing antibody structures inform therapeutic strategies. *Nature*. 588, 682–687 (2020).

136. D. Corti, D. Pinto, Y.-J. Park, M. Beltramello, A. C. Walls, M. A. Tortorici, S. Bianchi, S. Jaconi, K. Culap, F. Zatta, A. D. Marco, A. Peter, B. Guarino, R. Spreafico, E. Cameroni, J. B. Case, R. E. Chen, C. Havenar-Daughton, G. Snell, A. Telenti, H. W. Virgin, A. Lanzavecchia, M. S. Diamond, K. Fink, D. Veesler, Structural and functional analysis of a potent sarbecovirus neutralizing antibody, 1–28 (2020).

137. M. Yuan, N. C. Wu, X. Zhu, C.-C. D. Lee, R. T. Y. So, H. Lv, C. K. P. Mok, I. A. Wilson, A highly conserved cryptic epitope in the receptor-binding domains of SARS-CoV-2 and SARS-CoV. *Science*, eabb7269-10 (2020).

138. M. Sakharkar, C. G. Rappazzo, W. F. Wieland-Alter, C.-L. Hsieh, D. Wrapp, E. S. Esterman, C. I. Kaku, A. Z. Wec, J. C. Geoghegan, J. S. McLellan, R. I. Connor, P. F. Wright, L. M. Walker, Prolonged evolution of the human B cell response to SARS-CoV-2 infection. *Sci Immunol*. 6, e-abg6916 (2021).

139. C. Gaebler, Z. Wang, J. C. C. Lorenzi, F. Muecksch, S. Finkin, M. Tokuyama, A. Cho, M. Jankovic, D. Schaefer-Babajew, T. Y. Oliveira, M. Cipolla, C. Viant, C. O. Barnes, Y. Bram, G. Breton, T. Hägglöf, P. Mendoza, A. Hurley, M. Turroja, K. Gordon, K. G. Millard, V. Ramos, F. Schmidt, Y. Weisblum, D. Jha, M. Tankelevich, G. Martinez-Delgado, J. Yee, R. Patel, J. Dizon, C. Unson-O'Brien, I. Shimeliovich, D. F. Robbiani, Z. Zhao, A. Gazumyan, R. E. Schwartz, T. Hatziloannou, P. J. Bjorkman, S. Mehandru, P. D. Bieniasz, M. Caskey, M. C. Nussenzweig, Evolution of antibody immunity to SARS-CoV-2. *Nature*. 591, 639–644 (2021).

140. C. Atyeo, S. Fischinger, T. Zohar, M. D. Stein, J. Burke, C. Loos, D. J. McCulloch, K. L. Newman, C. Wolf, J. Yu, K. Shuey, J. Feldman, B. M. Hauser, T. Caradonna, A. Schmidt, T. J.

Suscovich, C. Linde, Y. Cai, D. Barouch, E. T. Ryan, R. C. Charles, D. Lauffenburger, H. Chu, G. Alter, Distinct early serological signatures track with SARS-CoV-2 survival. *Immunity*. 53, 524-532.e4 (2020).

141. N. C. Kyriakidis, A. López-Cortés, E. V. González, A. B. Grimaldos, E. O. Prado, SARS-CoV-2 vaccines strategies: a comprehensive review of phase 3 candidates. *Npj Vaccines*. 6, 28 (2021).

142. N. Pardi, M. J. Hogan, F. W. Porter, D. Weissman, mRNA vaccines — a new era in vaccinology. *Nat Rev Drug Discov*. 17, 261–279 (2018).

143. L. R. Baden, H. M. E. Sahly, B. Essink, K. Kotloff, S. Frey, R. Novak, D. Diemert, S. A. Spector, N. Rouphael, C. B. Creech, J. McGettigan, S. Khetan, N. Segall, J. Solis, A. Brosz, C. Fierro, H. Schwartz, K. Neuzil, L. Corey, P. Gilbert, H. Janes, D. Follmann, M. Marovich, J. Mascola, L. Polakowski, J. Ledgerwood, B. S. Graham, H. Bennett, R. Pajon, C. Knightly, B. Leav, W. Deng, H. Zhou, S. Han, M. Ivarsson, J. Miller, T. Zaks, C. S. Group, Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine. *New Engl J Med*. 384, 403–416 (2020).

144. F. P. Polack, S. J. Thomas, N. Kitchin, J. Absalon, A. Gurtman, S. Lockhart, J. L. Perez, G. P. Marc, E. D. Moreira, C. Zerbini, R. Bailey, K. A. Swanson, S. Roychoudhury, K. Koury, P. Li, W. V. Kalina, D. Cooper, R. W. Frenck, L. L. Hammitt, Ö. Türeci, H. Nell, A. Schaefer, S. Ünal, D. B. Tresnan, S. Mather, P. R. Dormitzer, U. Şahin, K. U. Jansen, W. C. Gruber, C. C. T. Group, Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine. *New Engl J Med* (2020), doi:10.1056/nejmoa2034577.

145. J. Sadoff, G. Gray, A. Vandebosch, V. Cárdenas, G. Shukarev, B. Grinsztejn, P. A. Goepfert, C. Truyers, H. Fennema, B. Spiessens, K. Offergeld, G. Scheper, K. L. Taylor, M. L. Robb, J. Treanor, D. H. Barouch, J. Stoddard, M. F. Ryser, M. A. Marovich, K. M. Neuzil, L. Corey, N. Cauwenberghs, T. Tanner, K. Hardt, J. Ruiz-Guiñazú, M. L. Gars, H. Schuitemaker, J. V. Hoof, F. Struyf, M. Douoguih, E. S. Group, Safety and Efficacy of Single-Dose Ad26.COV2.S Vaccine against Covid-19. *New Engl J Med* (2021), doi:10.1056/nejmoa2101544.

146. D. Y. Logunov, I. V. Dolzhikova, D. V. Shchegolyakov, A. I. Tukhvatulin, O. V. Zubkova, A. S. Dzharullaeva, A. V. Kovyrshina, N. L. Lubenets, D. M. Grousova, A. S. Erokhova, A. G. Botikov, F. M. Izhaeva, O. Popova, T. A. Ozharovskaya, I. B. Esmagambetov, I. A. Favorskaya, D. I. Zrelkin, D. V. Voronina, D. N. Shcherbinin, A. S. Semikhin, Y. V. Simakova, E. A. Tokarskaya, D. A. Egorova, M. M. Shmarov, N. A. Nikitenko, V. A. Gushchin, E. A. Smolyarchuk, S. K. Zyryanov, S. V. Borisevich, B. S. Naroditsky, A. L. Gintsburg, G.-C.-V. V. T. Group, Safety and efficacy of an rAd26 and rAd5 vector-based heterologous prime-boost COVID-19 vaccine: an interim analysis of a randomised controlled phase 3 trial in Russia. *Lancet.* **397**, 671–681 (2021).

**Chapter 2: Construction, characterization, and immunization of nanoparticles that display  
a diverse array of influenza HA trimers**

This Chapter describes the design and production of mosaic nanoparticles that co-display up to eight different hemagglutinins trimers derived from group 1 and group 2 HAs, as an immunization strategy to elicit a broadly reactive immune response against HA. The mosaic-HA particles were characterized through biochemical and structural methods. Immunizations with mosaic-HA particles were carried out and antibody as well as B-cell responses were characterized. Although immunization with these particles did induce robust antibody responses against the antigens presents, a benefit was not observed when compared to immunization with a mixture of homotypic particles

My contribution to this work was as the lead researcher of the project. I conceived and designed the study, prepared the reagents, and analyzed the data, and wrote the paper.

This work was published as

**Cohen, A.A.**, Yang, Z., Gnanapragasam, P.N.P., Ou, S., Dam, K.-M.A., Wang, H., and Bjorkman, P.J. (2021). Construction, characterization, and immunization of nanoparticles that display a diverse array of influenza HA trimers. *Plos One* 16, e0247963. 10.1371/journal.pone.0247963

## Abstract

Current influenza vaccines do not elicit broadly protective immune responses against multiple strains. New strategies to focus the humoral immune response to conserved regions on influenza antigens are therefore required for recognition by broadly neutralizing antibodies. It has been suggested that B-cells with receptors that recognize conserved epitopes would be preferentially stimulated through avidity effects by mosaic particles presenting multiple forms of a variable antigen. We adapted SpyCatcher-based platforms, AP205 virus-like particles (VLPs), and mi3 nanoparticles (NPs) to covalently co-display SpyTagged hemagglutinin (HA) trimers from group 1 and group 2 influenza A strains. Here we show successful homotypic and heterotypic conjugation of up to 8 different HA trimers to both VLPs and NPs. We characterized the HA-VLPs and HA-NPs by cryo-electron tomography to derive the average number of conjugated HAs and their separation distances on particles, and compared immunizations of mosaic and homotypic particles in wild-type mice. Both types of HA particles elicited strong antibody responses, but the mosaic particles did not consistently elicit broader immune responses than mixtures of homotypic particles. We conclude that covalent attachment of HAs from currently-circulating influenza strains represents a viable alternative to current annual influenza vaccine strategies, but in the absence of further modifications is unlikely to represent a method for making a universal influenza vaccine.

**Key words:** Influenza virus, Hemagglutinin nanoparticles, Hemagglutinin VLPs, SpyCatcher, Nanoparticle vaccine, Immunizations, Cryo-electron tomography

## Introduction

Each year, influenza virus infections affect 5-30% of the global population, resulting in millions of severe infections and hundreds of thousands of deaths [1]. Yearly epidemics are typically caused by the influenza type A, with a smaller number of infections resulting from type B. Vaccines can minimize the incidence of severe infections; however, they do not offer complete protection and have to be re-administered annually [1, 2]. The lack of complete efficacy of current vaccines can be attributed to several reasons. Mainly, the virus undergoes antigenic drift in which mutations accumulate over time that can allow the virus to evade the humoral immune response [1]. This requires that the vaccine formulation be renewed yearly so that the vaccine strains match the circulating strains as closely as possible. Influenza virus also features high antigenic diversity resulting in predominantly strain-specific antibody responses and making it difficult to recognize conserved regions on the viral antigens [2, 3]. Furthermore, through the mechanism of antigenic shift, the RNA segments from strains of different origins can reassort, resulting in new strains that are typically the cause of global pandemics that can rapidly circulate within an antigenically-naïve population [1]. Altogether, these necessitate the need for a universal flu vaccine that could confer protection against a broad swath of antigenically-distinct strains, thereby eliminating the need for yearly vaccines and offering protection against emergent pandemics.

The antibody response to influenza is primarily directed against the hemagglutinin (HA) and neuraminidase glycoproteins, which appear as a dense array of spikes on the surface of the viral particles [3, 4]. The majority of the neutralizing antibody response is against HA, the most abundant viral surface glycoprotein and the sialic acid-binding receptor that mediates fusion between the viral and

host membranes. Influenza HA is a trimer of HA1 and HA2 heterodimers, which can be subdivided into head and stalk domains [3, 4]. The HA head is composed of the middle portion of the HA1 sequence, contains the sialic acid binding site responsible for host cell recognition, and features high variability between different strains/subtypes. The HA2 subunit along with N- and C-terminal regions of HA1 encodes for the more conserved stalk domain, which contains the fusion peptide involved in viral/host cell membrane fusion [3, 4]. Antibodies against the immunodominant HA head can be strongly neutralizing, but are also strain specific, with the exception of antibodies that recognize the receptor binding site [3, 5, 6]. In contrast, the HA stem is immunosubdominant; however, stem antibodies are often broader, although generally less potent than anti-head antibodies, and can induce antibody-dependent cellular cytotoxicity (ADCC) responses [3, 7-9]. Within the past 10 years, broadly neutralizing antibodies (bNAbs) against influenza that target conserved HA stem epitopes have been discovered, but these antibodies have thus far been difficult to elicit [10]. It is generally believed that a broadly protective or “universal” vaccine would require the induction of anti-HA stem antibodies. As a result, there have been numerous attempts to refocus the immune response to these conserved epitopes [11-16].

One strategy to redirect the antibody response towards invariant epitopes was co-displaying influenza HAs from different strains on nanoparticles [16]. The rationale was to display HAs from several strains on a multimerized platform, such that any two adjacent HAs have a low probability of being identical, thereby giving a competitive advantage to B-cells with B-cell receptors (BCRs) that use avidity effects to recognize conserved epitopes shared between different strains. By contrast, BCRs that recognize strain-specific epitopes could not use avidity effects to bind adjacent HAs, thus would be less likely to be activated [16]. In this study, monomeric HA receptor binding domain (RBD)

sequences were fused to an engineered ferritin subunit to create self-assembling particles displaying up to eight different RBD sequences derived from H1N1 strains at 24 total positions [16]. The elicited humoral immune response in injected mice featured high breadth and potency against a panel of diverse H1N1 strains, which was most apparent the larger the number of HAs that were co-displayed, such that the simultaneous display of 8 different HAs elicited the greatest breadth in comparison with immunization of a cocktail of 8 different homotypic HA nanoparticles. Sorting and isolation of memory B-cells that were positive for HAs from two different strains further supported the use of this strategy for inducing cross-reactive B-cells. A logical follow up to this study would be to co-display HA ectodomain trimers including the stalk and head regions instead of RBD head domain monomers, with the hope of eliciting antibody lineages with increased breadth.

We reasoned that co-display of multiple HA trimers on a nanoparticle would be facilitated by a system in which soluble HA trimers could be covalently attached to a protein nanoparticle, thereby avoiding potential folding problems created by genetically fusing protomers from a trimer to a nanoparticle subunit. Numerous NP platforms and coupling strategies have been explored for vaccine design [17]. The “plug and display” strategy involves the use of virus-like particles (VLPs) or nanoparticles (NPs) fused to a SpyCatcher protein that is covalently conjugated to a purified antigen tagged with a short (13-residue) SpyTag [18, 19]. The conjugation involves the formation of an isopeptide bond between a lysine from the SpyCatcher protein and an aspartate from the SpyTag [20]. An advantage of the SpyCatcher-SpyTag system is that it allows for the spontaneous irreversible conjugation of a purified antigen with native-like post-translational modifications to a scaffold via an incubation of the antigen and scaffold proteins under physiological conditions. Available Spy-

Catcher protein scaffolds are highly versatile, coming in different forms that range from a bacteriophage AP205 T=3 icosahedral particle (180 SpyCatchers) to a designed dodecahedral NP called mi3 (60 SpyCatchers) [18, 19]. We recently used AP205 SpyCatcher-VLPs to display SpyTagged trimeric HIV-1 Env immunogens and demonstrated priming in immunized mice and non-human primates of B-cells carrying receptors displaying characteristics of V3-glycan patch-targeting HIV-1 bNAbs [21].

Here we describe the use of bacteriophage AP205-based SpyCatcher VLPs and engineered particle mi3-based SpyCatcher NPs [18, 19] to display a diverse array of HA ectodomain trimers from group 1 and group 2 influenza A strains. Successful conjugation was demonstrated by size-exclusion chromatography (SEC), SDS-PAGE, and electron microscopy (EM), with up to eight different HA trimers successfully conjugated to mosaic mi3 particles. Our results demonstrated that SpyCatcher-VLPs and SpyCatcher-NPs can be easily used to stably display at least 8 different trimeric antigens and that AP205-HA and mi3-HA particles produced strong immune responses in mice.

## Materials and methods

**Expression and purification of soluble HA trimers.** HA ectodomain trimers were expressed as shown schematically in Fig 1A with a C-terminal foldon trimerization domain, 13-residue SpyTag [20], and a 6x-His (modified from HA constructs in [22] to include a SpyTag). Genes corresponding to the modified HA1-HA2 sequences (residues 1-504 H3 numbering) from A/Aichi/02/1968 (Aichi; H3), A/Shanghai/1/2013 (SH13; H7), A/Jiangxi-Donghu/346/2013 JX346; H10), A/swine/HuBei/06/2009 (HB09; H4), A/California/04/2009 (CA09; H1), A/Vietnam/1203/ 2004 (Viet04; H5), A/Japan/305/1957 (JP57; H2), and A/guinea fowl/Hong Kong/1999 (WF10; H9N2)

were subcloned into a pTT5 expression vector. Genes encoding SpyTagged HAs with a Y98F mutation (H3 numbering) were constructed using site directed mutagenesis. HA ectodomain trimer constructs were expressed by transient transfection using the Expi293 Expression System (ThermoFisher), and soluble HA trimers were purified from transfected cell supernatants by standard Ni-NTA chromatography using a prepacked HisTrap™ HP column (GE Healthcare) and SEC using a HiLoad® 16/600 Superdex® 200 column (GE Healthcare). Proteins were concentrated using an Amicon Ultra 15 mL 30K concentrator (MilliporeSigma) and stored at 4°C in 20 mM Tris pH 8.0, 150 mM NaCl, 0.02% NaN<sub>3</sub> (TBS buffer).

HA ectodomain trimers for ELISAs were expressed as above without the 13-residue SpyTag or the Y98F substitution. Additional strains only used for ELISA include: A/shearwater/West Australia/2576/79 (WA79; H15) and A/flat-faced bat/Peru/033/2010 (Pe10; H18). The CA09-miniHA construct (construct #4900) [11] was subcloned into a pTT5 mammalian expression vector with a 6x-His tag and expressed and purified as described for the HA ectodomain trimers. For flow cytometry experiments, an Avi-tag was inserted after the C-termini of the Aichi, Viet04, and CA09 HAs with the Y98F substitution. Avi-tagged HAs were expressed and purified as described above and biotinylated using the Biotin ligase kit (Avidity) according to the manufacturer's protocol. Biotinylated CA09-HA, Aichi-HA and Viet04-HA were incubated with eBioscience™ Streptavidin APC, Streptavidin PE-eFluor™ 610, or Streptavidin PE (ThermoFisher) overnight at 4°C at a 1:1 molar ratio of HA trimer to streptavidin subunit.

**Expression of SpyCatcher-VLPs and SpyCatcher NPs.** pGEM-SpyCatcher-AP205-CP3 for expression of SpyCatcher-VLPs was the kind gift of Dr. Mark Howarth (Oxford University). pGEN

SpyCatcher AP205-CP3 was transformed into OverExpress™ C41(DE3) *E.coli* (Sigma). Single colonies were picked and inoculated into a 2xYT (Sigma) overnight starter culture and then grown in 1L 2xYT media with shaking at 220 rpm at 37°C until OD 0.5 ( $A_{600}$ ), after which they were induced with 0.42 mM IPTG and grown for 5 hours at 30°C. Cultures were harvested and pellets were frozen in lysis buffer (20mM Tris-HCl pH 7.8, 150mM NaCl, 0.1% Tween 20, 75 mM imidazole). For producing VLPs for conjugation, pellets were thawed and lysed with a cell disruptor in the presence of 2.0 mM PMSF (Sigma). The lysate was spun at 21,000xg for 30 min, filtered with a 0.2  $\mu$ m filter, and VLPs were isolated by Ni-NTA chromatography using a prepacked HisTrap™ HP column (GE Healthcare). SpyCatcher VLPs were eluted with 2.0 M imidazole, 50 mM glycine, 25 mM sodium citrate, 0.1% Tween 20, pH 8.5. Eluted VLPs were concentrated using an Amicon Ultra 15 mL 30K concentrator (MilliporeSigma) and further purified by SEC using a HiLoad® 16/600 Superdex® 200 (GE Healthcare) column equilibrated with 500 mM glycine pH 8.0, 250 mM sodium citrate, 1% Tween 20. VLPs were then stored at 4°C and used for up to 1 month for conjugations. SpyCatcher-VLPs precipitated out of solution over time and before conjugations they were either filtered with a 0.2  $\mu$ m filter or spun down at 21,000g for 10 min.

The pET28a His6-SpyCatcher-mi3 gene (Addgene) was transformed into BL21 (DE3)-RIPL *E.coli* (Agilent). Single colonies were picked and inoculated into an LB overnight starter culture, and grown in 1L LB media until OD 0.8 ( $A_{600\text{ nm}}$ ) with shaking at 220 rpm at 37°C, after which they were induced with 0.5 mM IPTG and grown for 16-20 hours at 20°C. Cultures were then harvested and pellets were frozen in lysis buffer (250 mM Tris-HCl pH 8.0, 150 mM NaCl, 50 mM imidazole, 0.02%  $\text{NaN}_3$ ). For producing NPs for conjugation, pellets were thawed and lysed with a cell disruptor in the presence of 2.0 mM PMSF (Sigma), and the lysate was spun at 21,000xg for 30 min, filtered with a 0.2

μm filter, and NPs were isolated by Ni-NTA chromatography using a prepacked HisTrap™ HP column (GE Healthcare), and eluting with 2.0 M imidazole, 20 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.02% NaN<sub>3</sub>. Eluted NPs were concentrated using an Amicon Ultra 15 mL 30K concentrator (MilliporeSigma) and further purified by SEC using a HiLoad® 16/600 Superdex® 200 (GE Healthcare) column equilibrated with 25 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.02% NaN<sub>3</sub>. NPs were then stored at 4°C and used for up to 1 month for conjugations. SpyCatcher-NPs precipitated out of solution over time, and before use for conjugations they were either filtered with a 0.2 μm filter or spun down at 21,000g for 10 min.

**Preparation of HA-VLPs and HA-NPs.** Purified SpyCatcher-VLPs or SpyCatcher-NPs were incubated with a 1.2-fold molar excess (HA protomer to VLP or NP subunit) of purified SpyTagged HA (either a single HA for making homotypic particles or an equimolar mixture of two or more HAs for making mosaic particles) at room temperature in TBS (25 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.02% NaN<sub>3</sub>) overnight. Conjugated VLPs or NPs were then separated from free HA trimers by SEC on a Superose 6 10/300 (GE Healthcare) column equilibrated with PBS (20 mM sodium phosphate pH 7.5, 150 mM NaCl). Fractions corresponding to conjugated VLPs or NPs were collected and analyzed via SDS-PAGE. Concentrations were determined using a Bio-Rad Protein Assay. For stability studies, mosaic NP preps were stored for a month at 4°C and then analyzed via SEC using a Superose 6 10/300 (GE Healthcare) column equilibrated with PBS (20 mM sodium phosphate pH 7.5, 150 mM NaCl).

**EM.** HA-conjugated and unconjugated VLPs and NPs were compared by negative-stain EM. Ultrathin, holey carbon-coated, 400 mesh Cu grids (Ted Pella, Inc.) were glow discharged for 60 s at

15 mA. A 3- $\mu$ l aliquot of SEC-purified HA-VLPs and HA-NPs diluted to approximately 40-100 ug/ml were applied to the grids for 60 s, and then negatively stained with 2% (w/v) uranyl acetate for 30 s. Data were collected with a FEI Tecnai T12 transmission electron microscope at 120 keV at 42,000x magnification.

SEC-purified HA-VLPs and HA-NPs were prepared on grids for cryo-ET using a Mark IV Vitrobot (ThermoFisher Scientific) operated at 21°C and 100% humidity. 3.1  $\mu$ L of sample was mixed with 1  $\mu$ L of 10 nm colloidal gold beads (Sigma-Aldrich) as fiducial markers and then applied to 300 mesh Quantifoil R2/2 grids, blotted for 3.5 s, and plunge-frozen in liquid ethane surrounded by liquid nitrogen. Cryo-ET was performed on a 300kV Titan Krios transmission electron microscope (ThermoFisher Scientific) equipped with a Gatan energy filter (slit width 20 eV) operating at a nominal 33,000x magnification. For HA-VLPs, tilt series were collected on a K2 direct electron detector (Gatan) with a pixel size of 2.23  $\text{\AA}\cdot\text{pixel}^{-1}$  using SerialEM software [25], a -3 to -6  $\mu$ m defocus range, and a total of 98  $e^-\cdot\text{\AA}^{-2}$  per tilt series. For HA-NPs, tilt series were recorded in counting mode on a K3 direct electron detector (Gatan) with a pixel size of 2.68  $\text{\AA}\cdot\text{pixel}^{-1}$  using SerialEM [25], a -4 to -5  $\mu$ m defocus range, and a total dose of  $\sim$ 140  $e^-\cdot\text{\AA}^{-2}$  per tilt series. For both data collections, tilt series images were collected using a dose-symmetric tilt scheme [26] ranging from -60° to 60° with 2° and 3° intervals for HA-VLPs and HA-NPs, respectively. Images were aligned and reconstructed using IMOD [27, 28].

**Immunizations.** All animal experiments were carried out in 4-6 week old female Balb/c mice obtained from Charles River Laboratories. The immunizations with HA-VLPs and HA-NPs in Fig 4A and Fig 5A, respectively, were done in Balb/c mice (n=4 in each group) through intraperitoneal (ip)

injections of 20  $\mu$ g of antigen in 200  $\mu$ L of 50% v/v of adjuvant (Sigma Adjuvant System®). For experiments in Fig 4A, mice were immunized on Day 0 and boosted on Day 14. Animals were bled weekly via tail veins. For animals in Fig 5A, mice were also boosted on Day 37. Mice were euthanized 2 weeks later (Day 49, 51), bled through cardiac puncture, and had their spleens harvested. For Fig 7A, mice (n=5 except for mi3- or VLP-immunized mice, where n=2) were immunized with the indicated immunogen in 100  $\mu$ L of 50% v/v of AddaVax™ adjuvant (Invivogen) and boosted with adjuvant on Day 14, 28, and 168. This adjuvant was chosen in order to compare with previous experiments [16, 29]. Mice were euthanized 2 weeks after the final boost (Day 182,183), bled through cardiac puncture, and had their spleens harvested. All blood samples were allowed to clot at room temperature in MiniCollect® Serum and Plasma Tubes (Greiner), and then serum was harvested, frozen in liquid nitrogen, and stored at -80°C until use. All of the animal experiments were performed using experimental protocols approved by the Institutional Animal Care and Use Committee (IACUC), California Institute of Technology (Protocol IA19-1725). Animals were euthanized at the end of the experiment and spleen tissue was harvested for in vitro analysis to look at the memory B-cell response. Animals were euthanized with CO<sub>2</sub> inhalation, confirmed by lack of heart-beat and/ respiratory rate.

**ELISAs.** Nunc® MaxiSorp™ 384-well plates (Sigma) were coated with 10  $\mu$ g/ml of a purified HA (without a SpyTag) in 0.1 M NaHC<sub>0</sub><sub>3</sub> pH 9.8 and stored overnight at 4°C. Plates were blocked with 3% bovine serum albumin (BSA) in TBS-T (TBS with 0.1% Tween 20) for 1 hr at room temperature. Plates were washed with TBS-T after each step. Serum was diluted 1:100 and then serially diluted

by 4-fold with TBS-T/3% BSA and added for 3 hr at room temperature. A 1:50,000 dilution of secondary HRP-conjugated goat anti-mouse IgG (Abcam) was added for 1 hr at room temperature. Plates were developed using SuperSignal™ ELISA Femto Maximum Sensitivity Substrate (ThermoFisher) and read at 425 nm. Curves were plotted and integrated to obtain the area under the curve (AUC) using Graphpad Prism 8.3. Statistical differences of AUC titers between groups were calculated using Tukey's multiple comparison test via Graphpad Prism 8.3.

**In vitro neutralization assays.** Neutralization assays were conducted using live PB1flank-eGFP virus for BSL 2 strains A/Aichi/02/1968 (X31; H3N2), A/California/04/2009 (CA09; H1N1), A/Texas/36/1991 (TX91; H1N1), and A/Wisconsin/67/2005 (WI05; H3N2) as described [30] using reagents kindly provided by Dr. Jesse Bloom (Fred Hutchinson). Plasma was set at a top dilution of 1:200 (for Fig 4C) or 1:100 (For Fig 5C) and serially diluted 5-fold (for Fig 4C) or 4-fold (For Fig 5C) for a total of 8 dilutions. Pseudovirus assays were conducted as described [31] for BSL 3 strains A/Shanghai/1/2013 (SH13; H7), A/Jiangxi-Donghu/346/2013 (JX346; H10), A/Vietnam/1203/ 2004 (Viet04-H5), and A/Netherlands/219/2003 (NL03-H7). Plasma was set at a top dilution of 1:200 (for Fig 4C) or 1:100 (For Fig 5C) and serially diluted 4-fold for a total of 8 dilutions. Neutralization data were plotted, curves were fit, and ID<sub>50</sub> values were calculated using Antibody Database [32]. Reported IC<sub>50</sub>s are geometric means, which are suitable for data sets covering multiple orders of magnitude [33]. Statistical differences of ID<sub>50</sub> titers between groups were calculated using Tukey's multiple comparison test via Graphpad Prism 8.3. Correlation between neutralization ID<sub>50</sub>s and ELISA AUC titers were calculated using the Pearson correlation function on Graphpad Prism 8.3.

**Flow cytometry.** Single cell suspensions were prepared from immunized mouse spleens by mechanical dissociation using the back of a syringe plunger. Cell suspensions in 70 µm cell strainers were washed in cold RPMI 1640 media and treated with ACK lysing buffer (Gibco) to lyse red blood cells. The resulting white blood cell preparation was resuspended in RPMI 1640 MACS and enriched for memory B-cells using the negative selection portion of the protocol in a mouse Memory B-cell Isolation Kit (Miltenyi). For the experiment in Fig 6A, enriched splenocytes were then stained with the following monoclonal antibodies and reagents: CD4-APC-eFluor 780 (clone: RM4-5), F4/80-APC-eFluor 780 (clone: BM8), CD8a-APC-eFluor 780 (clone: 53-6.7), Ly-6G-APC-eFluor 780 (clone: RB6-8C5), IgM-PerCP-eFluor 710 (clone: II/41) (eBioscience), CD19-FITC (clone: 6D5) (Biolegend), IgG1 BV421 (clone: X40), IgG2 BV421 (clone: R19-15) (BD Bioscience), and CA09-HA-APC, Aichi-HA-PEeflour610 and Viet04-HA-PE (prepared as described above). Cell viability was analyzed with Ghost Dye™ Violet 510 (Tonbo). For the experiment in Fig 8A, enriched splenocytes were stained with the following monoclonal antibodies and reagents: CD4-APC-eFluor 780 (clone: RM4-5), F4/80-APC-eFluor 780 (clone: BM8), CD8a-APC-eFluor 780 (clone: 53-6.7), Ly-6G-APC-eFluor 780 (clone: RB6-8C5), IgM- APC-eFluor 780 (clone: II/41) (Thermo-fisher Scientific), CD19-FITC (clone: 6D5) (Biolegend), IgG1 BV421 (clone: X40), IgG2 BV421 (clone: R19-15) (BD Bioscience), and CA09-HA-APC, Viet04-HA-PE for looking at CA09+Viet04+ B-cells, CA09-HA-APC and Aichi-HA-PE for looking at CA09+Aichi+ B-cells, or Sh13-HA-APC and Aichi-HA-PE for looking at Sh13+Aichi+ B-cells (all HA probes were prepared as described above). Cell viability was analyzed with Ghost Dye™ Violet 510 (Tonbo). Splenocytes were incubated for 30 min at 4°C in the dark and then washed twice with staining buffer (HBSS, 50 mM HEPES pH 7.4, 2.5 mg/ml BSA, 50 µg/ml DNase, 1 mM MgCl<sub>2</sub>). Stained cells were then analyzed with a SY3200 Cell Sorter (Sony) configured to detect 9 fluorochromes. 500,000-1,000,000 events were collected per sample

and analyzed via FlowJo software (TreeStar). Statistical differences of antigen-specific B-cell populations between groups were calculated using Tukey's multiple comparison test via Graphpad Prism 8.3. Correlation between percentage of antigen-specific B-cells and ELISA AUC titers were calculated using the Pearson correlation function on Graphpad Prism 8.3.

## Results and discussion

**Construction of HA-VLPs and HA-NPs.** We adapted the AP205 SpyCatcher-VLP platform that we had previously used to conjugate a trimeric HIV-1 immunogen [21] as a way to increase the intrinsic immunogenicity of HA, mask undesired epitopes located at the bottom of the HA trimer, and attach different HA trimers to the same particle. We reasoned that the SpyCatcher-VLP platform could be used to display more than one HA by incubating with equimolar amounts of different SpyTagged HAs. Although the SpyTagged HAs would be conjugated at random to available SpyCatcher proteins, there should be no advantage for the conjugation of one HA over another since they all contained the same SpyTag.

We first expressed and purified SpyTagged soluble HA trimers derived from 8 different influenza strains from group 1 and group 2 influenza A viruses (Fig 1A). The constructs for each HA protomer contained HA1 and the HA2 ectodomain (residue 1-503 H3 numbering) linked to a C-terminal foldon trimerization domain, a SpyTag, and a 6x-His tag (Fig 1A). The HAs for the SpyCatcher-mi3 conjugations included the sialic acid binding knockout mutation Y98F (except for the SpyTagged HAs used to conjugate the SpyCatcher-VLPs). SpyTagged HAs including the Y98F substitution purified

from the supernatants of transiently-transfected mammalian cells were verified to form monodisperse and well-behaved trimers by SEC and SDS-PAGE (Fig 1B).

We used SpyCatcher-AP205 VLPs and SpyCatcher-mi3 NPs as conjugation platforms for multivalent display of HA trimers. AP205-SpyCatcher VLPs are icosahedral capsids (T=3 symmetry) with 180 copies total, therefore 180 SpyCatchers were available for conjugation (Fig 1C). AP205-SpyCatcher VLPs were expressed in *E. coli* and purified via Ni-NTA affinity chromatography followed by SEC (Fig 1D) [18]. The SpyCatcher-AP205-VLPs eluted near the void volume as a single monodisperse peak. SpyCatcher-mi3-NPs are an engineered dodecameric scaffold with 60 total subunits, and therefore 60 conjugation sites [19] (Fig 1E). The SpyCatcher-mi3s were also expressed in *E. coli*, purified by Ni-NTA affinity chromatography followed by SEC, and analyzed with reducing SDS-PAGE (Fig 1F).

Using the SpyCatcher-AP205 VLPs, we first evaluated coupling of two recombinant HAs: A/California/04/09 H1 (CA09-HA) and A/Aichi/02/1968 H3 (Aichi-HA) (chosen to represent two strains that would normally be present in an annual influenza vaccine [2, 3]). Conjugations of the AP205-SpyCatcher VLPs were carried out by room temperature incubation with CA09-HA, Aichi-HA, or an equimolar mixture of both HAs in a 1.2 molar excess to the VLPs (HA protomer to VLP subunit) to prepare CA09-, Aichi-, and mosaic-2 VLPs, respectively (Fig 2A). VLP-conjugated HA trimers were separated from free trimers by SEC (Fig 2B), and successful conjugation of the SpyCatcher VLPs to Aichi-HA, CA09-HA, and both HAs was verified by a shift in apparent molecular weight (from 75 kDa to 100 kDa) detected by SDS-PAGE for HAs conjugated to the VLP subunits (Fig 2B).

Because VLPs conjugated with more than two different HAs tended to precipitate out of solution, we switched to the SpyCatcher-mi3 NP platform, which is similarly immunogenic as the AP205 platform, but has been shown to exhibit improved yields, stability, and uniformity [19]. In addition, we modified the SpyTagged HAs to include a receptor binding site mutation, Y98F (H3 numbering), to abolish sialic acid binding [34] that could result in interactions of aggregation of HAs on neighboring particles. Starting with 8 HAs from influenza group 1 and group 2 strains (Fig 1A) with pandemic potential [35], we made mosaic-2, mosaic-4 and mosaic-8 NPs (each with an equal representation of group 1 and group 2 strains) and the 8 corresponding homotypic HA-conjugated NPs (Fig 2C). Homotypic and mosaic HA-mi3s were purified via SEC, and conjugation was verified by a shift in apparent molecular weight (from 75 kDa to >100 kDa) detected by SDS-PAGE for HAs conjugated to the mi3 subunits (Fig 2D,E). To assess stability of the conjugated NPs, mosaic NP samples stored for one month at 4°C were analyzed for degradation by SEC, revealing little to no free HA trimer for both the mosaic-4 and mosaic-8 NPs (Fig 2E).

**EM characterization of HA-VLPs and HA-NPs.** Negative-stain EM revealed increased diameters for conjugated VLPs and NPs compared with their unconjugated counterparts (Fig 3A). HA-conjugated VLPs were also examined by single-particle cryo-EM. 2D class averages of mosaic-2 VLPs showed ordered density for the AP205 VLP, but blurred densities for attached HAs (Fig 3B), suggesting variability in trimer orientations with respect to the VLP surface.

Since the HA trimer densities could not be reliably interpreted by single-particle cryo-EM, we used cryo-ET to derive 3D reconstructions of individual HA-conjugated VLPs and NPs. Tomograms (Supplemental Movies 1 and 2) showed particles with average diameters of 60 nm (HA-VLPs) and 50

nm (HA-NPs) and revealed densities for individual HA trimers on VLPs and NPs. The trimers were separated by distances of ~7-10 nm and ~12-15 nm for VLPs and NPs, respectively (measured between the head regions of trimer axes on adjacent HAs). To estimate the number of conjugated HA trimers, we counted HA densities in ~3 nm tomographic slices of individual HA-VLPs and HA-NPs at their widest diameters, where the symmetries of each type of particle predicted a maximum of 20 potential attachment sites. We found 9-16 HA densities for conjugated VLPs and 6-8 densities for conjugated NPs, corresponding to occupancies of 45-80% (VLPs) and 30-40% (NPs). Since AP205 VLPs and mi3 NPs contain 180 or 60 SpyCatcher domains, respectively, this translates to ~81-144 conjugated HA trimers per AP205 VLP and ~18-24 trimers per mi3 NP.

**Immunizations with homotypic- and mosaic-HA-VLPs/NPs.** Our next goal was to determine whether mosaic HA-VLPs induced a more cross-reactive humoral immune response compared with a mixture of the corresponding homotypic HA-VLPs. We first immunized one group of four mice with mosaic-2 VLPs (presenting CA09 plus Aichi HAs) and a second group of four mice with an equal mixture of CA09-VLPs and Aichi-VLPs (admix-2) (Fig 4A). In addition, we immunized groups of mice with only Aichi-VLPs or only CA09-VLPs. In all cases, mice were primed with equal doses of VLPs plus adjuvant, boosted 2 weeks later without adjuvant, and bled weekly for serum analyses. Serum ELISAs were performed to measure IgG binding to purified HAs from a panel of group 1 and group 2 influenza A strains (Fig 4B). As expected, IgG titers elicited by immunization with mosaic-2- and admix-2-immunized mice were similar to titers elicited by CA09-VLP immunization against CA09 HA and Aichi-VLP immunization against Aichi-HA. Against heterotypic HAs not coupled to the VLPs (Viet04, Jp57, WF10, Sh13, and JX346 HAs), IgG titers were similar for both mosaic-2- and admix-2-immunized mice, although titers were consistently higher compared with both CA09-

VLP- and Aichi-VLP-immunized mice. Thus in terms of elicited IgG binding of HAs, it appeared that immunizing with the mosaic-2 VLPs that contained group 1 and group 2 HAs was no better at inducing cross-reactive binding of HAs from divergent strains than the corresponding admixture. However, the mosaic-2 and admix-2 injections induced heterologous breadth that could not be explained by the overlapping immunogenicity of the homotypic VLPs.

Next, we determined neutralizing activity of the serum samples using *in vitro* neutralization assays (using infectious viruses for BSL 2 strains and pseudoviruses for BSL 3 strains) against a panel of group 1 and group 2 influenza A strains (Fig 4C). For the mosaic-2- and admix-2-immunized mice, neutralizing titers against homotypic infectious virus strains (CA09 and Aichi) were consistent with the ELISA titers against these strains. Against the Viet04 and JX346 pseudoviruses, neutralization titers were not detectable except for one animal in the admix-2 group. Against the Sh13 pseudovirus, neutralizing titers for the mosaic-2-immunized mice were higher than for the other groups, although the spread in potency was broad and overlapped with the other groups. When considering the neutralization and ELISA results together, the mosaic-2 VLPs did not induce greater breadth than the corresponding mixture of homotypic VLPs.

In order to determine whether mosaic HA-NPs with higher valencies could elicit antibody responses with higher breadth, we conducted experiments similar to those described for VLPs to compare injections of mosaic-2, -4 and -8 NPs with the corresponding admixtures of homotypic NPs (Fig 5A), a CA09-NP homotypic control, and unconjugated SpyCatcher-NPs. A final boost was performed 5 weeks after the first prime and animals were sacrificed 2 weeks later to harvest spleens for B-cell analysis. Serum IgG titers from Day 21 and Day 49 were measured by ELISA against a panel of

purified HAs from homotypic and heterotypic group1 and group 2 strains (Fig 5B) and against unconjugated SpyCatcher-NPs (mi3 NPs in Fig 5A). All groups of mice exhibited antibody responses to unconjugated SpyCatcher-mi3, suggesting that SpyCatcher and/or NP epitopes are accessible on HA-conjugated particles.

Against HAs from homotypic strains (CA09 and Aichi), the serum from mice immunized with both the admix and mosaic NPs featured equivalent titers, with the exception of the admix-4 mouse group, which responded with overall lower titers and included 2 mice with no responses. Against the Viet04 and Sh13 HAs (presented on particles with valencies of 4 and 8), the response was slightly higher for the mosaic-2 NPs than for the CA09-NP and admix-2 groups, whereas the mosaic-4, mosaic-8 and admix-8 groups showed higher titers, although over a broad range that included mice with poor responses even against strains of HA that were presented on admix-4 NPs. Against the Jp57, JX346, WF10, and HB09 HAs (only present on the valency-8 NPs), the admix-8 and mosaic-8 titers were equivalent to each other and the highest on average. The titers against heterologous HAs from the CA09-NP-injected mice were similar to titers from mice injected with unconjugated SpyCatcher-NPs, with the exception of responses against Jp57 HA. Responses to the mosaic-2 NPs were slightly higher than responses against admix-2 and CA09-NP, again except for the recognition of Jp57 HA. One animal in the mosaic-4 group exhibited high titers against HAs from all strains in comparison to the admix-4 mice, with the mosaic-4 responses being on par with responses from the animals immunized with the valency-8 NPs. Finally, against HAs from Pe10 and WA79 (not presented on the any of the NPs), serum titers were low for most of the injected animals except for some animals from the mosaic-4 group (e.g., the animal that exhibited high titers against strains not represented on the mosaic-4 NPs), and admix-8- and mosaic-8-immunized mice.

ELISAs were also used to evaluate recognition of CA09-miniHA, a stabilized stem-only construct derived from CA09 HA [11] to investigate whether there was preferred recognition of stem epitopes by the animals immunized with mosaic NPs. We found that the serum response against the CA09 stem was equivalent to responses against the head-containing CA09 HA trimer, with the CA09-NP-immunized mice exhibiting the highest titers (Fig 5B), suggesting that the mosaic NPs did not preferentially elicit anti-stem responses.

ELISA titers determined for serum samples obtained on Day 49 showed similar responses as Day 21 titers with a few exceptions. For example, the mosaic-8 NP-immunized mice had lower titers compared to the admix-8 mice against some of the HAs, suggesting that the additional immunizations were not consistently resulting in strong immune responses. The mice immunized with the mosaic-4 NPs mounted more robust responses, although two of the animals exhibited low titers against all of the strains (as compared with three animals from Day 21). Based on ELISAs, we did not find strong evidence of increased cross-reactivity induced in animals immunized with the mosaic NPs compared with animals injected with the corresponding admix-NPs of the same valency. However, for the mosaic-4 NP group, one animal repeatedly showed high titers of IgG binding to HAs from all strains tested, suggesting that this animal may have induced cross-reactive antibodies.

We also conducted *in vitro* neutralization assays using Day 49 serum against a panel group 1 and group 2 influenza strains (Fig 5C). For CA09, neutralization titers correlated with ELISA titers (S1A Fig), with CA09-NP-immunized mice showing the highest neutralization titers. The neutralizing response to Aichi HA also correlated with ELISA titers (S1C Fig). For the two heterotypic H1 and H3

HAs (TX91 and WI05 strains), serum from all animals was non-neutralizing, suggesting that neutralizing antibodies that cross-react within the H1 and H3 subtypes were not induced. For Viet04 and Sh13, neutralization correlated with the ELISA titers (S1B and S1D Fig) with valency 4 and 8 mosaic and admix particles showing higher neutralization titers as expected. Against NL03, the neutralizing responses were difficult to interpret due to high background neutralization from the unconjugated mi3 control serum. For JX346, neutralization titers also correlated with ELISA results (S1E Fig). Overall, it appeared that the mosaic NPs did not offer an advantage compared to corresponding mixtures of homotypic particles in induction of neutralizing antibodies, although the mosaic-4 groups included some animals in which greater breadth was induced than admix-4 animals.

**B-cell responses induced by mosaic- versus homotypic-NP immunizations.** In order to determine if cross-reactive B-cells were elicited in mice immunized with the mosaic NPs, IgG+ B-cells from immunized mouse spleens were probed for binding to soluble HAs from three different influenza strains using flow cytometry (Fig 6A). The percent binding was determined by gating antigen-specific populations to compare populations positive for CA09, Viet04, or Aichi HAs alone, and for double-positive populations representing B-cells that exhibited cross-reactivity (Fig 6B). As expected, the CA09+ population was the largest for the CA09-NP-immunized mice, with the rest of the mosaic and admix groups eliciting lower proportions of the CA09+ B-cells. Admix and mosaic groups with valencies of 4 and 8 elicited a similar level of Viet04+ IgG+ B-cells, as expected since Viet04 HA was present only on these NPs. Interestingly, the mosaic-2-immunized mice elicited a somewhat lower, but detectable, number of Viet04+ B-cells, which were not present in the admix-2 and homotypic CA09-NP samples. Except for the CA09-NP-immunized mice, all animals showed

Aichi+ B-cells; however, the mosaic-2-immunized mice elicited the largest number of antigen-specific B-cells, consistent with ELISA and neutralization results (Fig 6A-B). Interestingly, very few CA09+/Viet04+ B-cells were induced in all immunized animals, with the exception of one animal in the mosaic-4-immunized group, which also featured a high serum IgG and neutralizing response. The CA09-NP and admix-8 immunized mice also induced double-positive B-cells, although to a lesser extent. CA09+/Aichi+ and Viet04+/Aichi+ double-positive B-cells were not detected for any of the animals (data not shown). As expected, antigen-specific B-cell populations (CA09+, Viet04+ and Aichi+) correlated strongly with ELISA serum binding (Aichi-HA, Viet04-HA, and Aichi-HA, respectively; S2A-S2C Fig). Interestingly, the percent of double-positive CA09+Viet04+ B-cells correlated with serum titers for Pe10-HA (S2D Fig), a strain not represented on any of the particles. This suggests that animals that induced CA09+Viet04+ B-cells also had cross-reactive serum antibodies. Although there was no significant difference in the induction of double-positive B-cells between the mosaic versus admix NP-immunized mice, there is some capacity for the HA-conjugated NPs to induce cross-reactive B-cells and antibodies.

**Comparison of Mosaic NP and VLP immunizations.** Several possibilities could account for why no significant differences between mosaic NPs and the corresponding admixture of homotypic NPs were observed. One reason is that there were some animals from each group that did not respond strongly to either prime or boost as determined by ELISA (Fig 5B). Another possibility is that the mi3 NP platform is not as immunogenic as the AP205 VLP platform, which can serve as a self-adjuvant via toll-like receptors since AP205 carries bacterial nucleic acid [18, 19].

Another animal experiment was conducted to compare mosaic VLPs and mosaic NPs with their counterpart admixtures. Mosaic VLPs and mosaic NPs were prepared with valencies of -4 and -8 along with the corresponding admixtures of homotypic VLPs and NPs (S2A and S2B Fig). Groups of 5 mice were immunized with mosaic and admixture VLPs and NPs, CA09-NP, and CA09-VLP, as well as unconjugated NPs and VLPs as controls (Fig 7A). Mice were then boosted with the same antigen in the presence of adjuvant a total of 3 times over the course of 4 months. Mice were bled every two weeks after each immunization. Two weeks after the third boost, mice were sacrificed for B-cell analysis using harvested spleens.

Serum IgG titers were measured via ELISAs using samples from Day 28 against HAs from a panel of group 1 and group 2 strains (Fig 7B). There was no significant difference between antibody titers elicited by mosaic NPs or mosaic VLPs and their counterpart admixtures of equivalent valency against any of the strains tested. Furthermore, there was no major difference between antibody titers from animals immunized with mosaic VLPs versus mosaic NPs. Serum IgG responses against both SpyCatcher-mi3 NPs and SpyCatcher-AP205 VLPs were high not significantly different between all animal groups, suggesting a strong background response against both spycatcher and mi3 or AP205 platforms (S4 Fig).

However, there was a statistically significant difference in ELISA titers against CA09-H1, when comparing mice immunized with CA09-VLPs with respect to mice immunized with mosaic-4 VLPs ( $p=0.0018$ ) and mosaic-8 VLPs ( $p=0.0084$ ). The difference between CA09 mi3 versus the mosaic NPs was not significant. This suggests that the increase of valency of strains represent on the mosaic VLPs reduced the humoral response against at least one of the component strains.

Somewhat surprisingly, mice immunized with either CA09-NP or CA09-VLP, which only presented the group 1 CA09 H1N1 HA, elicited antibody titers that were cross-reactive against all of the strains tested including the group 2 HAs, often to a similar level as the admix and mosaic VLPs and mosaic NPs that had those strains represented (Fig 7B), suggesting that immunization with a monovalent particle can be sufficient to induce cross-reactivity. Day 42 ELISA titers showed a similar trend (S2B Fig).

**Flow cytometry comparisons of mosaic NP and VLP immunizations.** The antigen-specific B-cell response for the mice immunized with mosaic VLPs or mosaic NPs was characterized using flow cytometry (Fig 8A). IgG+ splenocytes were analyzed for binding to a panel of soluble HAs derived from four strains: two from group 1 (CA09 and Viet04) and two from group 2 (Aichi and Sh13). The percent binding was determined for each antigen by gating the antigen-specific CD19+, IgG+, B-cell population that recognized either CA09, Viet04, Sh13, and Aichi HAs alone (Fig 8A), or that recognized double-positive populations that represented cross-reactive B-cells (Fig 8B). Three sets of double-positive cross-reactive antigen specificities were interrogated: CA09+/Viet04+ to look for group 1 breadth, Sh13+/Aichi+ to look for group 2 breadth, and CA09+/Aichi+ to look for group 1/group 2 breadth (Fig 8B).

Single positive populations for CA09, Viet04, Aichi, or Sh13 correlated with the Day 28 ELISA titers (S5A-S5D Fig). CA09-NP and CA09-VLP elicited a significantly higher percentage of CA09+ B-cells with respect to both the mosaic and admix versions of NPs and VLPs (Fig 8A), in agreement with ELISA results in which the CA09-VLPs elicited higher anti-CA09 titers than mosaic-4 and mosaic-8

VLPs (Fig 7B). There were no statistical significant differences in the percent of antigen-specific B-cells between mosaic-4 and mosaic-8 NPs and VLPs. However, as a general trend, the higher the mosaic valency, the lower the percent of strain-specific B-cells that were elicited, especially in the case of Viet04+ and Aichi+ B-cells (Fig 8A). Interestingly, mice immunized with CA09-NP and CA09-VLP induced antigen-specific B-cells that were specific to Viet04, Aichi, and Sh13 (Fig 8A). This could explain why cross-reactive ELISA titers were observed for these animals against HAs from every strain that was tested.

Similar to the results shown in Fig 6B, induction of cross-reactive CA09+/Viet04+ B-cells were rare (Fig 8B). The difference in the percent of CA09+/Viet04+ B-cells between mosaic, admix, and homotypic VLPs/NPs was therefore not significant. Interestingly, both CA09-NP and Ca09-VLPs were able to induce CA09+/Viet04+ B-cells, suggesting that immunization with monovalent CA09-VLPs/NPs was sufficient to induce cross-reactive B-cells. As previously observed, the percent of CA09+/Viet04+ B-cells correlated with the Day 28 serum ELISA titers against Pe10, a mismatched strain not represented on any of the VLPs or NPs (S5E Fig,  $p=0.0234$ ),

Induction of Ca09+/Aichi+ B-cells was observed, although rarely, making it unclear whether they represented B-cells that were cross-reactive to group 1 and group 2 HAs (Fig 8B). Sh13+/Aichi+ B-cells were also observed at a low frequency (Fig 8B). Since both of these populations were rare, there was no significant difference in the percent of double-positive B-cells between each group of immunized mice.

## Conclusions

Attempts to develop broadly protective influenza vaccines have been challenging partly due to the immunodominance hierarchy of antibody epitopes on HA. The variable epitopes on the HA head tend to be more easily recognized than invariant stem epitopes, therefore driving a predominantly strain-specific immune response [2, 3]. A potential strategy to redirect the antibody response towards more conserved stem epitopes is to co-display influenza HAs from different antigenically-distinct strains on particles. A previous study demonstrated the potential for this approach in that antibody responses with greater breadth were observed for mice injected with mosaic HA receptor binding domain particles compared with counterpart homotypic admixtures [16]. Here we sought to extend these results by preparing homotypic and mosaic particles containing trimeric HAs that included stem epitopes that are not present on monomeric HA receptor binding domains. Since HA trimers cannot be fused to ferritin nanoparticles, as previously done to prepare the monomeric HA receptor binding domain particles [16], we used a “plug and display” strategy [18] to covalently couple trimeric HAs to symmetric particles with different numbers of attachment sites (VLPs with 180 attachment sites and NPs with 60 attachment sites), thereby developing a simple method to make homotypic particles displaying a single strain of HA and mosaic particles displaying HAs derived from up to 8 strains. We demonstrated successful conjugation of HA trimers using biochemical methods and EM imaging, including cryo-ET to examine coupling densities of HA on VLPs and NPs. Our biochemical and EM analyses of HA-VLP and HA-NP particles provide useful characterizations for future efforts to utilize the SpyCatcher-SpyTag “Plug and Display” approach [18] for homotypic and heterotypic display of oligomeric antigens.

Our results showed that immunizations with mosaic particles conjugated with HA trimers did not offer a clear advantage in the induction of cross-reactive B-cells compared with immunization of

mixtures of homotypic particles. The finding that mosaic particles conjugated with monomeric HA receptor binding domains showed increased induction of cross-reactive B cells compared with admixtures [16], but that mosaic HA trimer particles compared with admix HA trimer particles did not, suggests that potential advantages of mosaic presentation may be related to particular forms of an antigen. For example, the monomeric HA antigens coupled to ferritin were limited to inducing HA head-specific antibodies [16], whereas the trimeric HA ectodomains contained head epitopes as well as stem epitopes that may be partially occluded from interactions with BCRs. In addition, the HA head monomers that were coupled to ferritin were restricted to the H1N1 subfamily [16], whereas our study involved HA trimers derived from group 1 and group 2 influenza strains. Consistent with our results, a recent study using designed nanoparticles to present trimeric HAs from influenza A and B strains also reported no increased breadth of antibody responses against mosaic particles compared with admixtures [29].

Although a clear difference in the degree of cross-reactive B-cells induced by mosaic NPs versus admix NPs when presenting a trimeric HA antigen has not yet been demonstrated, we observed cross-reactive B-cells in response to injections of mosaic and admixture particles. In particular, VLPs and NPs including CA09 HA induced broad responses for both homotypic and heterotypic particles. This suggests the inclusion of CA09 HA antigens on particles in future vaccines. In addition, although mosaic particles and the corresponding admixtures of homotypic particles induced similar levels of increased breadth, the use of mosaic NPs presents a potential therapeutic advantage, i.e., production of a mosaic NP would require purification of one set of particles, whereas use of admix NPs of the same valency would require purification and then mixing of multiple parti-

cles prior to immunization. Thus mosaic particles presenting HA antigens derived from multiple influenza strains should be considered as a potential vaccine strategy, and the SpyCatcher-SpyTag “Plug and Display” system [18] can be used to quickly combine different mixtures of oligomeric antigens for preparation of mosaic particles.

## Acknowledgements

We thank Mark Howarth (Oxford University) for providing plasmids and advice for VLP expression and purification, Jesse Bloom (Fred Hutchinson) for reagents for the infectious virus neutralization assays, Jost Vielmetter and Pauline Hoffmann at the Caltech Beckman Institute Protein Expression Center for help with protein production, Andrey Malyutin and Songye Chen (Caltech) for help with cryo-EM data collection, Rochelle Diamond and Jamie Tijerina at the Caltech Flow Cytometry/Cell Sorting Facility for help in the flow cytometry experiments and analysis, Jennifer Keeffe for implementation of influenza neutralization assays, Claudia Jette for help with figure preparation, Harry Gristick for VLP and NP images for figures, and Andrew Flyak, Jennifer Keeffe, and Claudia Jette for critical reading of the manuscript. EM was done in the Beckman Institute Resource Center for Transmission Electron Microscopy at Caltech. This work was supported by the National Institute Of Allergy And Infectious Diseases of the National Institutes of Health 1R01AI129784 (PJB) and the National Institutes of Health Grant P50 AI150464 (PJB).

## References

- 1.Krammer F, Smith GJD, Fouchier RAM, Peiris M, Kedzierska K, Doherty PC, Palese P, Shaw ML, Treanor J, Webster RG, Garcia-Sastre A (2018) Influenza. *Nat Rev Dis Primers* 4(1):3.
- 2.Sautto GA, Kirchenbaum GA, Ross TM (2018) Towards a universal influenza vaccine: different approaches for one goal. *Virology journal* 15(1):17.
- 3.Krammer F (2019) The human antibody response to influenza A virus infection and vaccination. *Nat Rev Immunol* 19(6):383-97.
- 4.Air GM (2015) Influenza virus antigenicity and broadly neutralizing epitopes. *Current opinion in virology* 11:113-21.
- 5.Angeletti D, Gibbs JS, Angel M, Kosik I, Hickman HD, Frank GM, Das SR, Wheatley AK, Prabhakaran M, Leggat DJ, McDermott AB, Yewdell JW (2017) Defining B cell immunodominance to viruses. *Nat Immunol* 18(4):456-63.
- 6.Schmidt AG, Therkelsen MD, Stewart S, Kepler TB, Liao HX, Moody MA, Haynes BF, Harrison SC (2015) Viral receptor-binding site antibodies with diverse germline origins. *Cell* 161(5):1026-34.
- 7.Ekiert DC, Bhabha G, Elsliger MA, Friesen RH, Jongeneelen M, Throsby M, Goudsmit J, Wilson IA (2009) Antibody recognition of a highly conserved influenza virus epitope. *Science* 324(5924):246-51.
- 8.Joyce MG, Wheatley AK, Thomas PV, Chuang GY, Soto C, Bailer RT, Druz A, Georgiev IS, Gillespie RA, Kanekiyo M, Kong WP, Leung K, Narpala SN, Prabhakaran MS, Yang ES, Zhang B, Zhang Y, Asokan M, Boyington JC, Bylund T, Darko S, Lees CR, Ransier A, Shen CH, Wang L, Whittle JR, Wu X, Yassine HM, Santos C, Matsuoka Y, Tsybovsky Y, Baxa U, Program NCS, Mullikin JC, Subbarao K, Douek DC, Graham BS, Koup RA, Ledgerwood JE, Roederer M, Shapiro L, Kwong PD, Mascola JR, McDermott AB (2016) Vaccine-Induced Antibodies that Neutralize Group 1 and Group 2 Influenza A Viruses. *Cell* 166(3):609-23.
- 9.Lang S, Xie J, Zhu X, Wu NC, Lerner RA, Wilson IA (2017) Antibody 27F3 Broadly Targets Influenza A Group 1 and 2 Hemagglutinins through a Further Variation in VH1-69 Antibody Orientation on the HA Stem. *Cell reports* 20(12):2935-43.
- 10.Tan HX, Jegaskanda S, Juno JA, Esterbauer R, Wong J, Kelly HG, Liu Y, Tilmanis D, Hurt AC, Yewdell JW, Kent SJ, Wheatley AK (2019) Subdominance and poor intrinsic immunogenicity limit humoral immunity targeting influenza HA stem. *J Clin Invest* 129(2):850-62.
- 11.Impagliazzo A, Milder F, Kuipers H, Wagner MV, Zhu X, Hoffman RM, van Meersbergen R, Huijzingh J, Wanningen P, Verspuij J, de Man M, Ding Z, Apetri A, Kukrer B, Sneekes-Vriese E, Tomkiewicz D, Laursen NS, Lee PS, Zakrzewska A, Dekking L, Tolboom J, Tettero L, van Meerten S, Yu W, Koudstaal W, Goudsmit J, Ward AB, Meijberg W, Wilson IA, Radosevic K (2015) A stable trimeric influenza hemagglutinin stem as a broadly protective immunogen. *Science* 349(6254):1301-6.
- 12.Yassine HM, Boyington JC, McTamney PM, Wei CJ, Kanekiyo M, Kong WP, Gallagher JR, Wang L, Zhang Y, Joyce MG, Lingwood D, Moin SM, Andersen H, Okuno Y, Rao SS, Harris AK, Kwong PD, Mascola JR, Nabel GJ, Graham BS (2015) Hemagglutinin-stem nanoparticles generate heterosubtypic influenza protection. *Nat Med* 21(9):1065-70.

13.Nachbagauer R, Liu WC, Choi A, Wohlbold TJ, Atlas T, Rajendran M, Solorzano A, Berlanda-Scorza F, Garcia-Sastre A, Palese P, Albrecht RA, Krammer F (2017) A universal influenza virus vaccine candidate confers protection against pandemic H1N1 infection in preclinical ferret studies. *NPJ Vaccines* 2:26.

14.Liu WC, Nachbagauer R, Stadlbauer D, Solorzano A, Berlanda-Scorza F, Garcia-Sastre A, Palese P, Krammer F, Albrecht RA (2019) Sequential Immunization With Live-Attenuated Chimeric Hemagglutinin-Based Vaccines Confers Heterosubtypic Immunity Against Influenza A Viruses in a Preclinical Ferret Model. *Front Immunol* 10:756.

15.Weidenbacher PA, Kim PS (2019) Protect, modify, deprotect (PMD): A strategy for creating vaccines to elicit antibodies targeting a specific epitope. *Proc Natl Acad Sci U S A* 116(20):9947-52.

16.Kanekiyo M, Joyce MG, Gillespie RA, Gallagher JR, Andrews SF, Yassine HM, Wheatley AK, Fisher BE, Ambrozak DR, Creanga A, Leung K, Yang ES, Boyoglu-Barnum S, Georgiev IS, Tsybovsky Y, Prabhakaran MS, Andersen H, Kong WP, Baxa U, Zephir KL, Ledgerwood JE, Koup RA, Kwong PD, Harris AK, McDermott AB, Mascola JR, Graham BS (2019) Mosaic nanoparticle display of diverse influenza virus hemagglutinins elicits broad B cell responses. *Nat Immunol* 20(3):362-72.

17.Brune KD, Howarth M (2018) New Routes and Opportunities for Modular Construction of Particulate Vaccines: Stick, Click, and Glue. *Front Immunol* 9:1432.

18.Brune KD, Leneghan DB, Brian IJ, Ishizuka AS, Bachmann MF, Draper SJ, Biswas S, Howarth M (2016) Plug-and-Display: decoration of Virus-Like Particles via isopeptide bonds for modular immunization. *Scientific reports* 6:19234.

19.Bruun TUJ, Andersson AC, Draper SJ, Howarth M (2018) Engineering a Rugged Nanoscaffold To Enhance Plug-and-Display Vaccination. *ACS Nano* 12(9):8855-66.

20.Zakeri B, Fierer JO, Celik E, Chittock EC, Schwarz-Linek U, Moy VT, Howarth M (2012) Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin. *Proc Natl Acad Sci U S A* 109(12):E690-7.

21.Escolano A, Gristick HB, Abernathy ME, Merkenschlager J, Gautam R, Oliveira TY, Pai J, West AP, Jr., Barnes CO, Cohen AA, Wang H, Golijanin J, Yost D, Keeffe JR, Wang Z, Zhao P, Yao KH, Bauer J, Nogueira L, Gao H, Voll AV, Montefiori DC, Seaman MS, Gazumyan A, Silva M, McGuire AT, Stamatatos L, Irvine DJ, Wells L, Martin MA, Bjorkman PJ, Nussenzweig MC (2019) Immunization expands B cells specific to HIV-1 V3 glycan in mice and macaques. *Nature* 570(7762):468-73.

22.Ekiert DC, Friesen RH, Bhabha G, Kwaks T, Jongeneelen M, Yu W, Ophorst C, Cox F, Korse HJ, Brandenburg B, Vogels R, Brakenhoff JP, Kompier R, Koldijk MH, Cornelissen LA, Poon LL, Peiris M, Koudstaal W, Wilson IA, Goudsmit J (2011) A highly conserved neutralizing epitope on group 2 influenza A viruses. *Science* 333(6044):843-50.

23.Shishovs M, Rummieks J, Diebolder C, Jaudzems K, Andreas LB, Stanek J, Kazaks A, Kotelovica S, Akopjana I, Pintacuda G, Koning RI, Tars K (2016) Structure of AP205 Coat Protein Reveals Circular Permutation in ssRNA Bacteriophages. *J Mol Biol* 428(21):4267-79.

24.Hsia Y, Bale JB, Gonon S, Shi D, Sheffler W, Fong KK, Nattermann U, Xu C, Huang PS, Ravichandran R, Yi S, Davis TN, Gonon T, King NP, Baker D (2016) Design of a hyperstable 60-subunit protein dodecahedron. [corrected]. *Nature* 535(7610):136-9.

25.Mastronarde DN (2005) Automated electron microscope tomography using robust prediction of specimen movements. *J Struct Biol* 152(1):36-51.

26.Hagen WJH, Wan W, Briggs JAG (2017) Implementation of a cryo-electron tomography tilt-scheme optimized for high resolution subtomogram averaging. *J Struct Biol* 197(2):191-8.

27.Kremer JR, Mastronarde DN, McIntosh JR (1996) Computer visualization of three-dimensional image data using IMOD. *J Struct Biol* 116(1):71-6.

28.Mastronarde DN, Held SR (2017) Automated tilt series alignment and tomographic reconstruction in IMOD. *J Struct Biol* 197(2):102-13.

29.Boyoglu-Barnum S, Ellis D, Gillespie RA, Hutchinson GB, Park Y-J, Moin SM, Acton O, Ravichandran R, Murphy M, Pettie D, Matheson N, Carter L, Creanga A, Watson MJ, Kephart S, Vaile JR, Ueda G, Crank MC, Stewart L, Lee KK, Guttman M, Baker D, Mascola JR, Veesler D, Graham BS, King NP, Kanekiyo M (2020) Elicitation of broadly protective immunity to influenza by multivalent hemagglutinin nanoparticle vaccines. *bioRxiv*.

30.Bloom JD, Gong LI, Baltimore D (2010) Permissive secondary mutations enable the evolution of influenza oseltamivir resistance. *Science* 328(5983):1272-5.

31.Temperton NJ, Hoschler K, Major D, Nicolson C, Manvell R, Hien VM, Ha do Q, de Jong M, Zambon M, Takeuchi Y, Weiss RA (2007) A sensitive retroviral pseudotype assay for influenza H5N1-neutralizing antibodies. *Influenza Other Respir Viruses* 1(3):105-12.

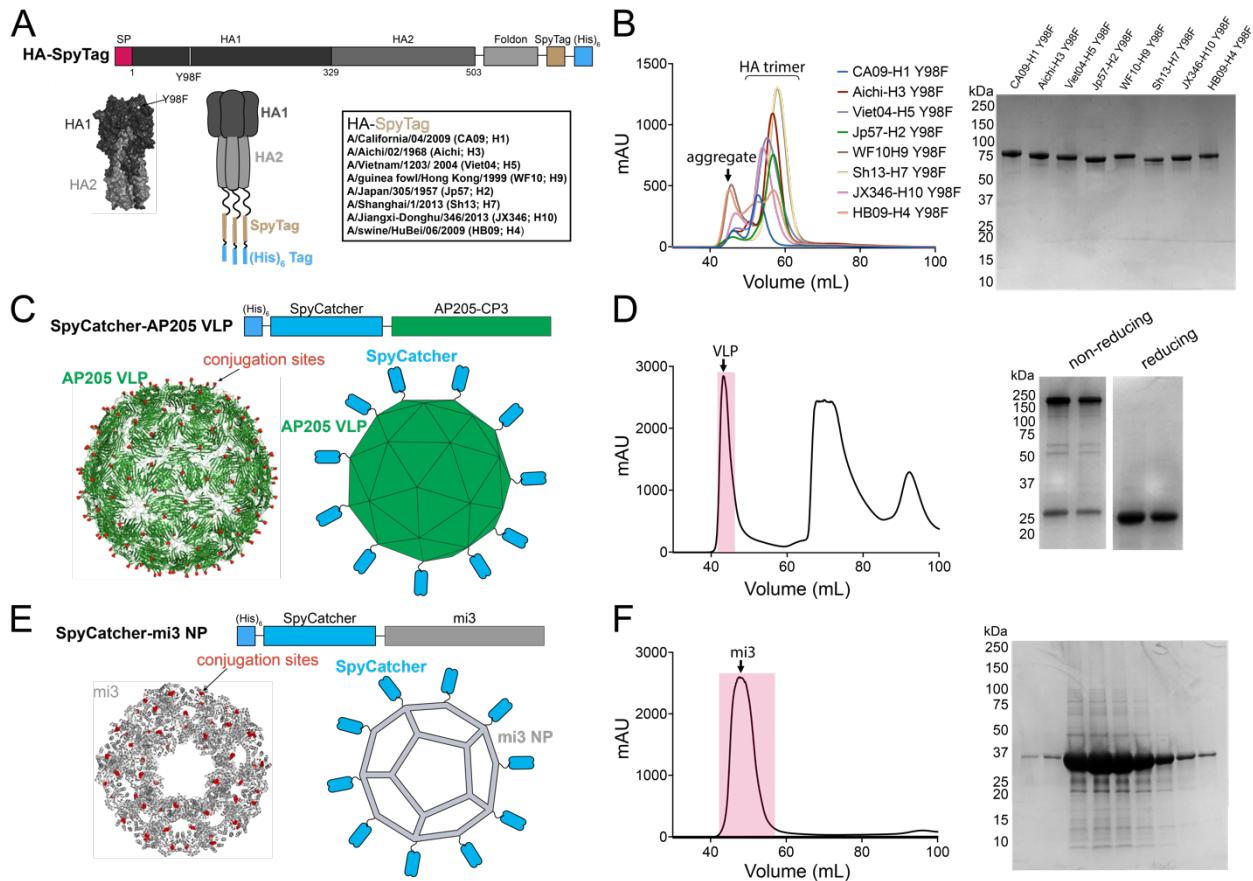
32.West AP, Jr., Scharf L, Horwitz J, Klein F, Nussenzweig MC, Bjorkman PJ (2013) Computational analysis of anti-HIV-1 antibody neutralization panel data to identify potential functional epitope residues. *Proc Natl Acad Sci U S A* 110(26):10598-603.

33.Sheskin D (2004) *Handbook of Parametric and Nonparametric Statistical Procedures*. 3rd ed. Boca Raton: Chapman & Hall/CRC. 1193 p.

34.Whittle JR, Wheatley AK, Wu L, Lingwood D, Kanekiyo M, Ma SS, Narpala SR, Yassine HM, Frank GM, Yewdell JW, Ledgerwood JE, Wei CJ, McDermott AB, Graham BS, Koup RA, Nabel GJ (2014) Flow cytometry reveals that H5N1 vaccination elicits cross-reactive stem-directed antibodies from multiple Ig heavy-chain lineages. *J Virol* 88(8):4047-57.

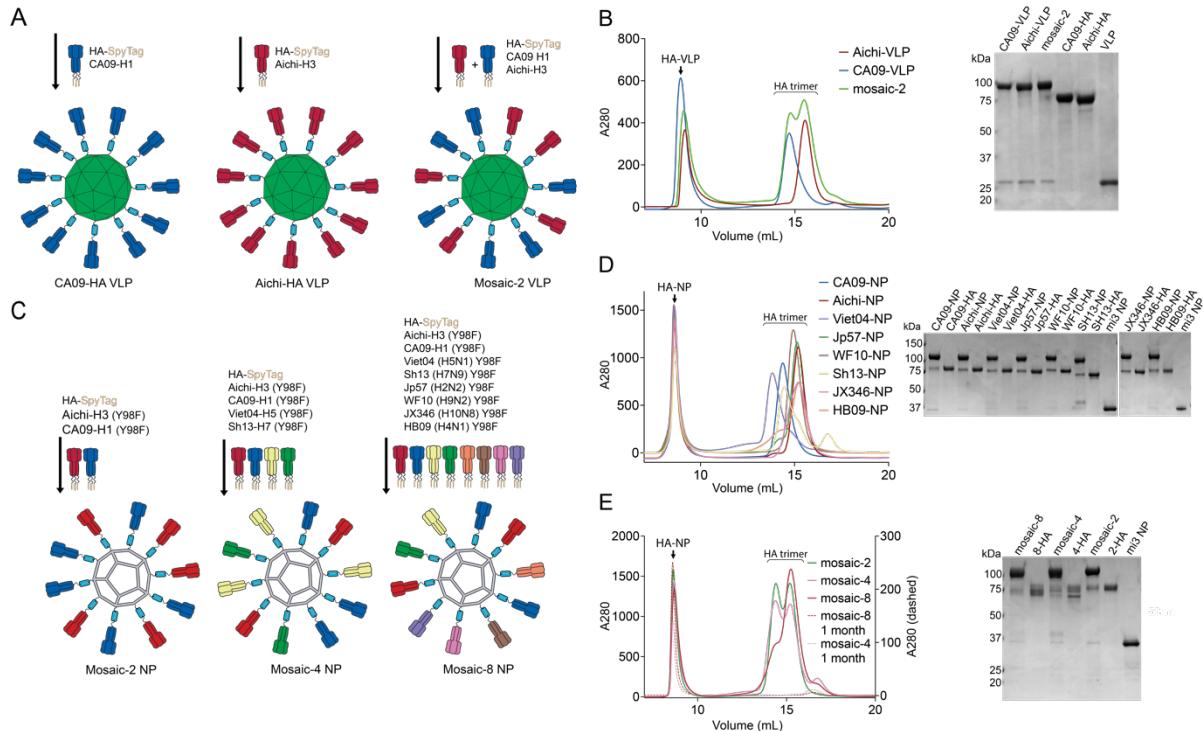
35.Mostafa A, Abdelwhab EM, Mettenleiter TC, Pleschka S (2018) Zoonotic Potential of Influenza A Viruses: A Comprehensive Overview. *Viruses* 10(9).

## Figure legends

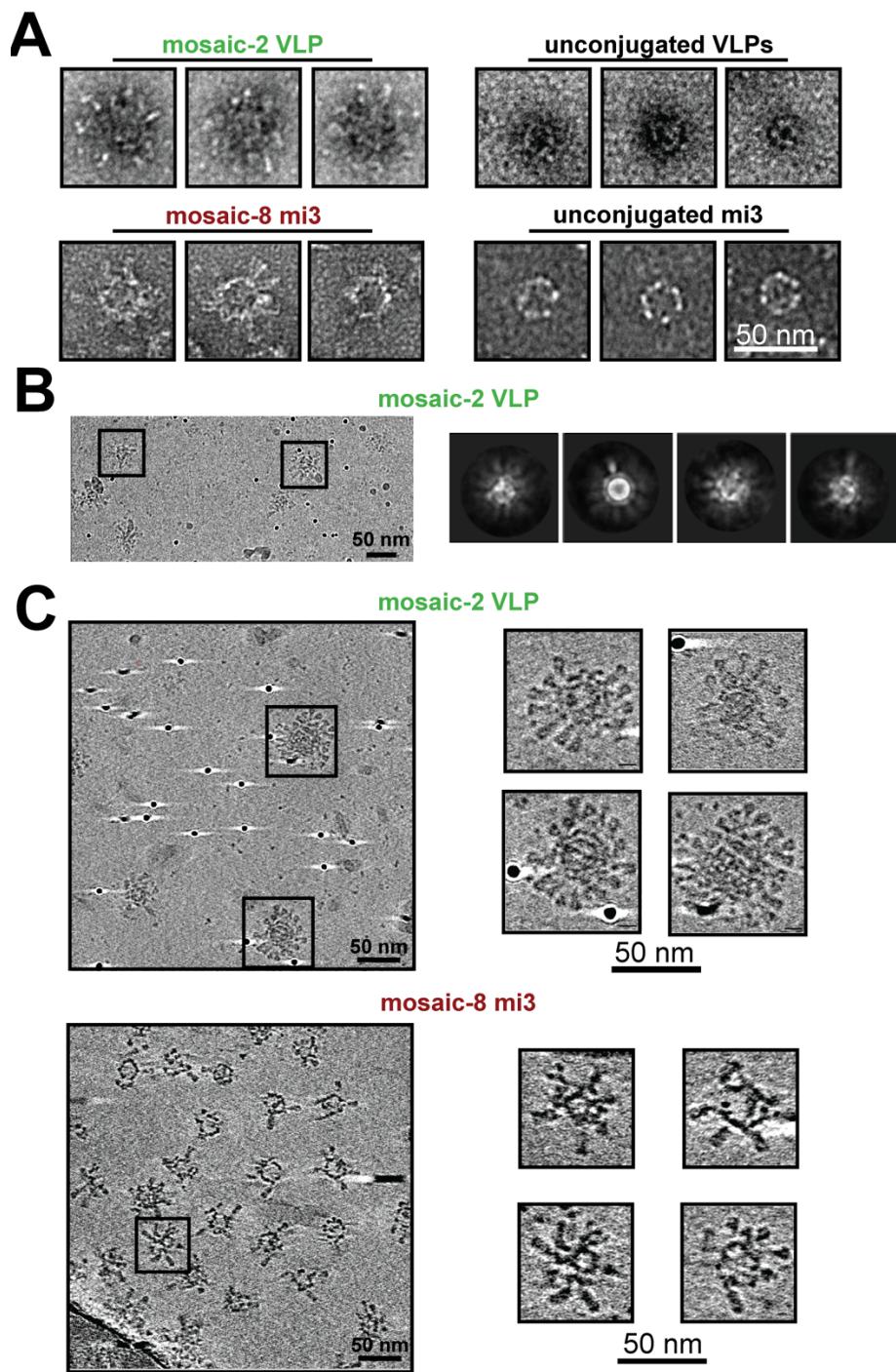


**Figure 1. Design and characterization of SpyTagged HAs, SpyCatcher VLPs, and SpyCatcher-NPs.** **A.** Top: Schematic of the SpyTagged HA construct (SP = signal peptide). The HA2 ectodomain is followed by a foldon trimerization domain from T4 fibritin, a 13-residue SpyTag, and a 6x-His tag. Amino acids are numbered according to the H3 nomenclature. Horizontal lines represent Gly<sub>4</sub>Ser linkers. Bottom: Surface representation of an HA trimer structure (PDB 3VUN), schematic of a SpyTagged HA, and list of influenza strains from which SpyTagged HAs were derived. **B.** SEC profiles and reducing SDS-PAGE analysis of 8 purified SpyTagged HAs that contain Y98F substitution. **C.** SpyCatcher-AP205 VLPs. Top: Schematic of construct. Bottom: EM structure of T=3 AP205 particle (PDB 5FS3) [23] with the locations of SpyCatcher fusion sites indicated by red dots (left) and schematic SpyCatcher-VLP (right). **D.** Purification of SpyCatcher-VLPs. Left: SEC profile with peak representing properly-assembled VLPs indicated. Right: Reducing and non-reducing SDS-PAGE of two fractions corresponding to the VLP fractions in red on the SEC trace. **E.** SpyCatcher-mi3 NPs. Top: schematic of construct. Bottom: Cryo-EM structure of I3-01 particle re-

lated to mi3 [24] with the locations of SpyCatcher fusion sites indicated by red dots (left) and schematic SpyCatcher-NP (right). **F.** Purification of SpyCatcher-NPs. Left: SEC profile. Right: Reducing SDS-PAGE of fractions corresponding to the mi3 fractions shaded in red on the SEC trace.

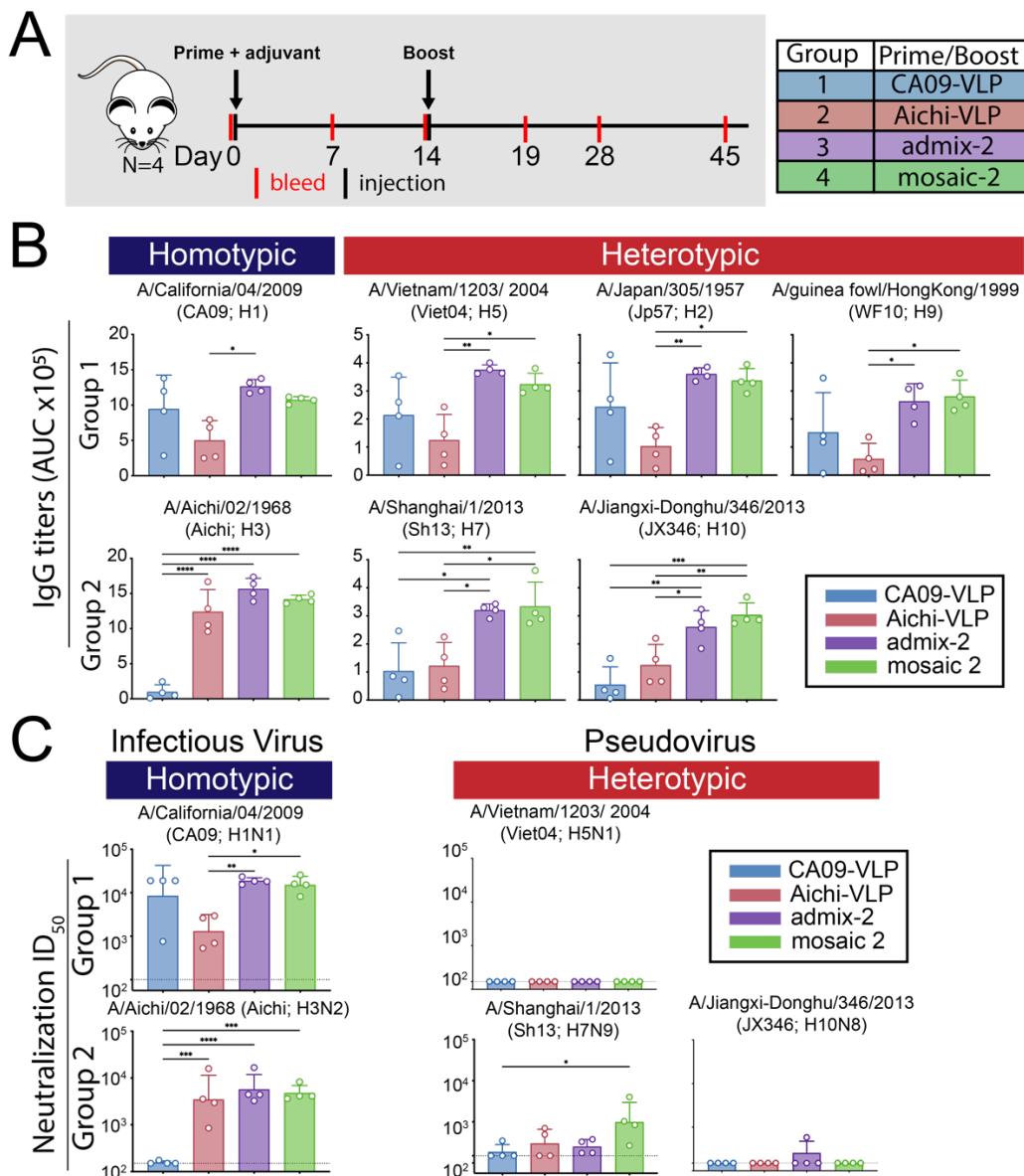


**Fig 2. Conjugation of SpyCatcher-VLPs and -NPs. A.** SpyCatcher-AP205-VLP conjugations with SpyTagged-HA trimers. **B.** Purification of conjugated SpyCatcher-VLPs. Left: SEC separation of conjugated VLPs from free HA trimers. Right: Reducing SDS-PAGE analysis of VLPs and purified HAs. **C.** SpyCatcher-mi3 NP conjugations with SpyTagged-HA Y98F trimers. **D.** Purification of homotypic SpyCatcher-NPs. Left: SEC separation of conjugated NPs from free HA trimers. Right: Reducing SDS-PAGE analysis of NPs and purified HAs. **E.** Purification of heterotypic mosaic NPs. Left: SEC separation of conjugated NPs from free HA trimers, including SEC profile of purified conjugated NPs after one month storage at 4 °C. Right: Reducing SDS-PAGE analysis of NPs and purified HAs.

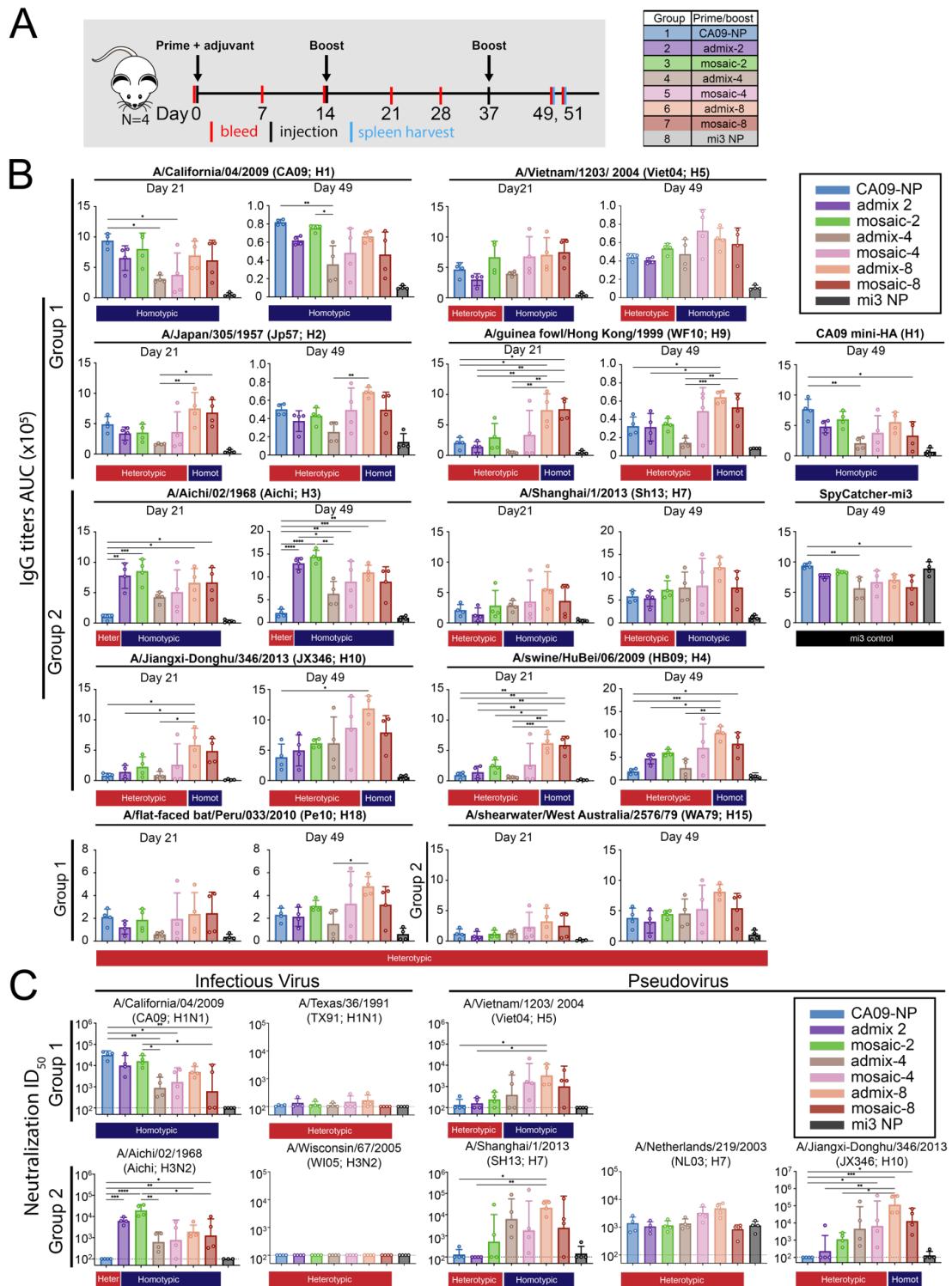


**Fig 3. EM of conjugated VLPs and NPs.** Scale bars shown apply to all images in each panel. **A.** Negative-stain EM of HA-conjugated VLPs and mi3 NPs compared with unconjugated counterparts.

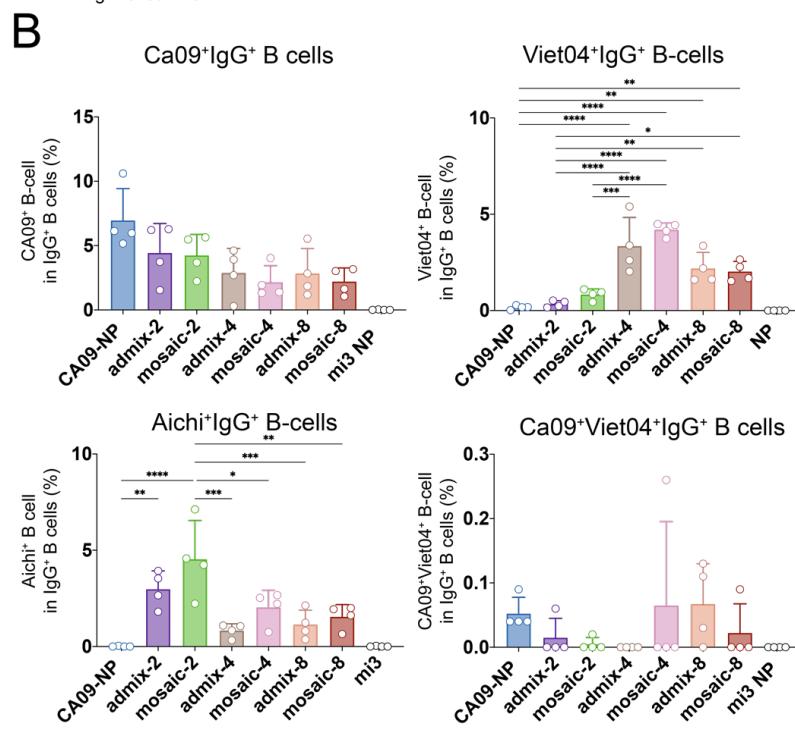
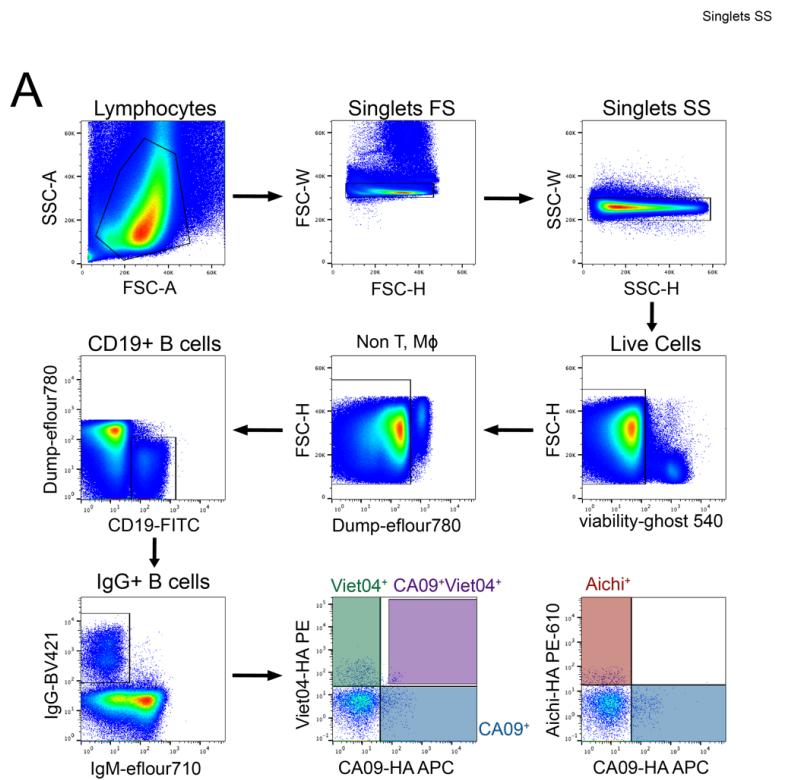
**B.** Cryo-EM micrograph of HA-VLP sample (left) and representative 2D class averages (right). Densities for HA trimers are blurry in the class averages, likely because the trimers occupy different positions on individual particles. **C.** Cryo-ET imaging of HA particles. Computationally-derived tomographic slices of HA-VLP (top panels; 2.78 nm slices) and HA-NP (bottom panels; 3.21 nm slices). Slices derived from the widest portions of representative particles are shown to the right in each panel.



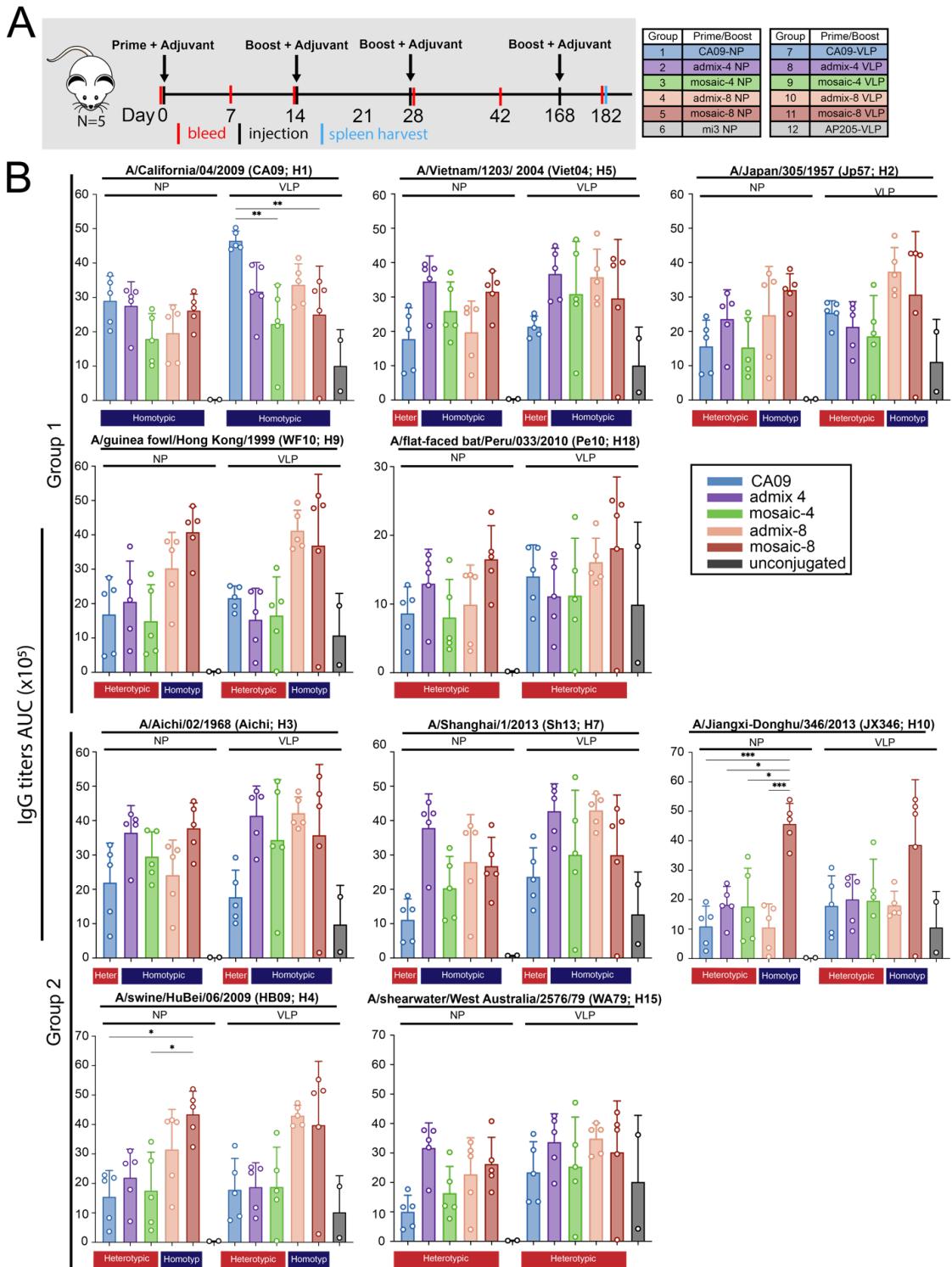
**Fig 4. Immunizations with HA-VLPs.** **A.** Schematic of the immunization protocol using HA-VLPs (wt HA). Four animals were used for each sample group. **B.** Serum antibody response to wt HA was measured by ELISA and shown as area under the curve (AUC) of Day 28 serum sample to group 1 and group 2 HA trimers. Each dot represents serum from one animal, with arithmetic means and standard deviations represented by rectangles and horizontal lines, respectively. Homotypic strains (present on the mosaic-2 VLP) and heterotypic strains (not present on the mosaic-2 VLP) are indicated by the blue and red rectangles, respectively, above the ELISA data. Significant differences between groups represented by horizontal lines are indicated by asterisks:  $p<0.05$  \*,  $p<0.01$  \*\*,  $p<0.001$  \*\*\*,  $p<0.0001$  \*\*\*\*. **C.** Serum neutralization titers from Day 45 determined by *in vitro* neutralization assays using infectious virus or pseudoviruses. Each dot represents serum from one animal, with geometric mean and geometric standard deviations represented by rectangles and horizontal lines, respectively. Dotted lines indicate limits of detection.



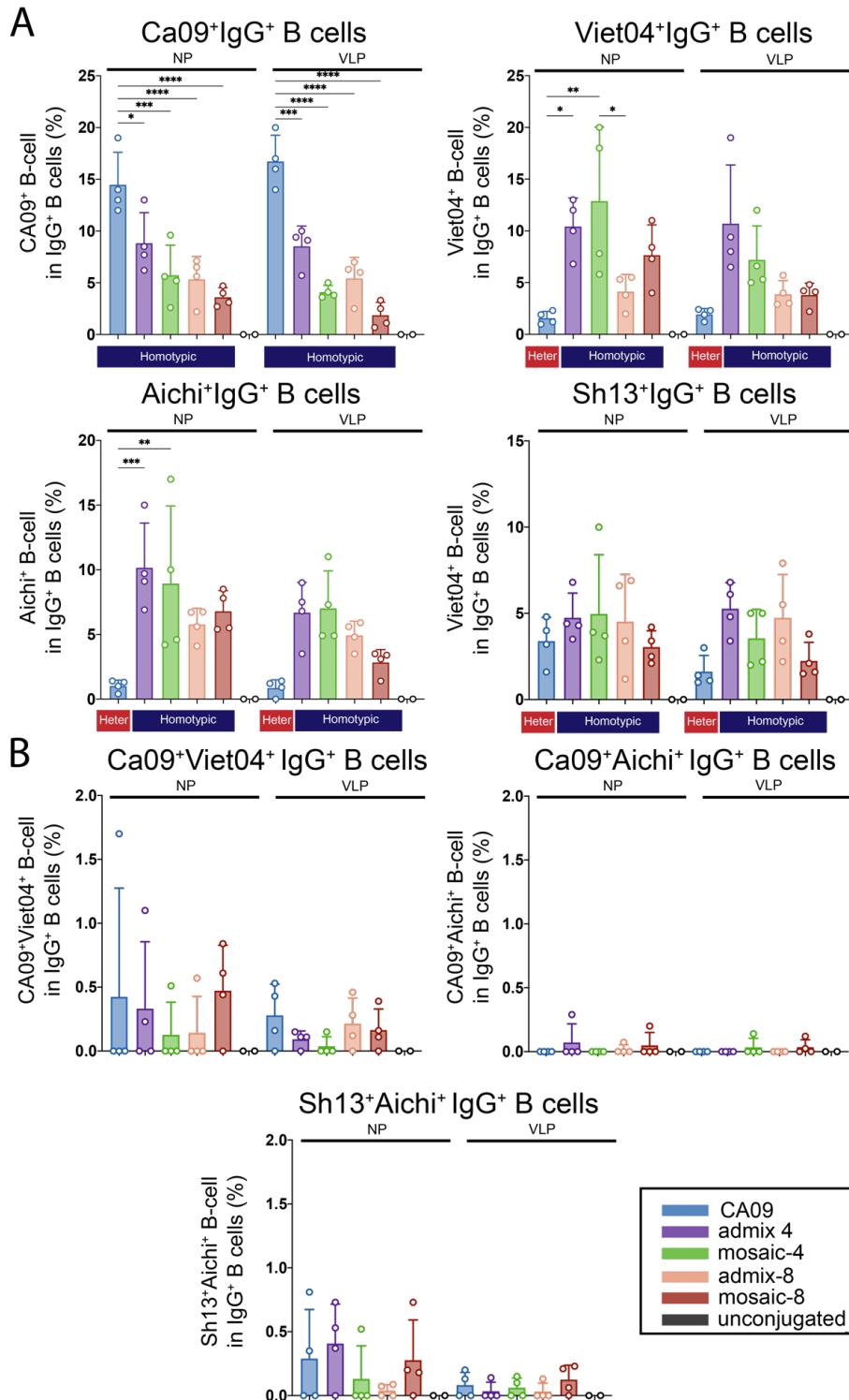
**Fig 5. Immunizations with HA-NPs.** **A.** Schematic of the immunization protocol using HA-NPs (Y98F). Four animals were used for each sample group. **B.** Serum antibody response to wt HA was tested by ELISA and shown as binding as area under the curve (AUC) of Day 21 and Day 49 serum to recombinant group 1 and group 2 HA trimers. Each dot represents serum from one animal, with means and standard deviations represented by rectangles and horizontal lines, respectively. Homotypic strains that were present on the mosaic NPs and heterotypic strains that were not present are indicated by the blue and red rectangles, respectively, above the ELISA data. Significant differences between groups represented by horizontal lines are indicated by asterisks:  $p<0.05$  \*,  $p<0.01$  \*\*,  $p<0.001$  \*\*\*,  $p<0.0001$  \*\*\*\*. **C.** Serum neutralization titers from Day 45 determined by *in vitro* neutralization assays using infectious virus or pseudoviruses. Each dot represents serum from one animal, with geometric means and geometric standard deviations represented by rectangles and horizontal lines, respectively. Dotted lines indicate limits of detection. ND = not determined.



**Fig 6. B-cell responses induced by mosaic NP immunizations. A.** Gating strategy for flow cytometry experiments using single cell suspension from spleens harvested from immunized mice in Fig 5A. Anti-CD3, anti-CD8, anti-F4/80, and anti-Ly6G were used to remove T cells, macrophages, monocytes, and neutrophils. Cells were then gated to isolate CD19/IgG-positive and IgM-negative B-cells, which were probed for binding to CA09-HA-APC (Y98F) (allophycocyanin), Aichi-HA-PE-eFluor610 (Y98F) (phycoerythrin-eFluor 610), and/or Viet04-HA-PE (phycoerythrin) (Y98F). **B.** Percentage of CA09+, Viet04+, Aichi+, and CA09+/Viet04+ in IgG+ B-cells plotted for each group. Significant differences between groups represented by horizontal lines are indicated by asterisks: : p<0.05 =\*, p<0.01=\*\*, p<0.001=\*\*\*, p<0.0001=\*\*\*\*. Differences with no significance are not shown, and significant differences between HA-NPs and mi3-NPs are also not shown

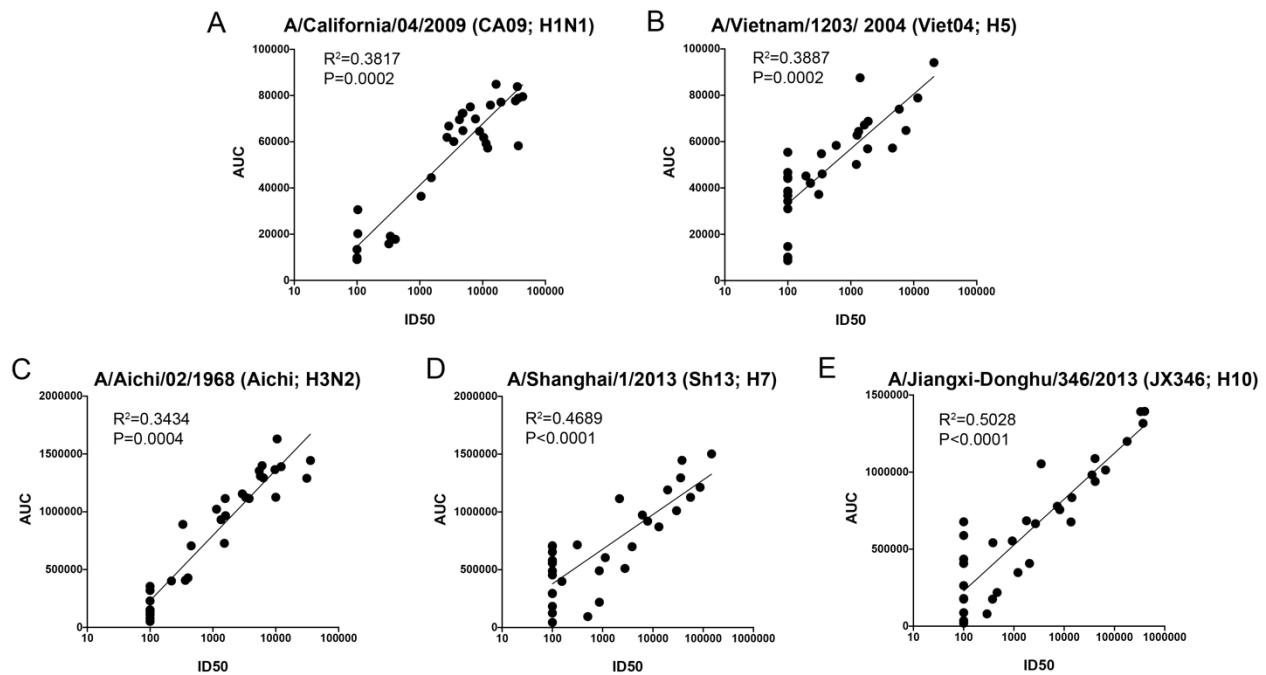


**Fig 7. Immunizations using Mosaic VLPs and NPs.** **A.** Schematic of the immunization protocol (five animals per injection) using HA-NPs (Y98F) and HA-VLPs (Y98F). Mice were immunized in the presence of Addavax adjuvant. **B.** Serum antibody response to wt HA shown by ELISA binding as area under the curve (AUC) of Day 28 serum to recombinant group 1 and group 2 HA trimers. Each dot represents serum from one animal, with means and standard deviations represented by rectangles and horizontal lines, respectively. Homotypic strains that were present on the mosaic NPs and heterotypic strains that were not present are indicated by the blue and red rectangles, respectively, above the ELISA data. Significant differences between groups represented by horizontal lines are indicated by asterisks:  $p<0.05 = ^*$ ,  $p<0.01=^{**}$ ,  $p<0.001=^{***}$ , and  $p<0.0001=^{****}$ . Differences with no significance are not shown, and significant differences between HA-NPs and mi3-NPs, HA-VLPs, and Ap205-VLPs, and HA-mi3s and HA-VLPs are also not shown.

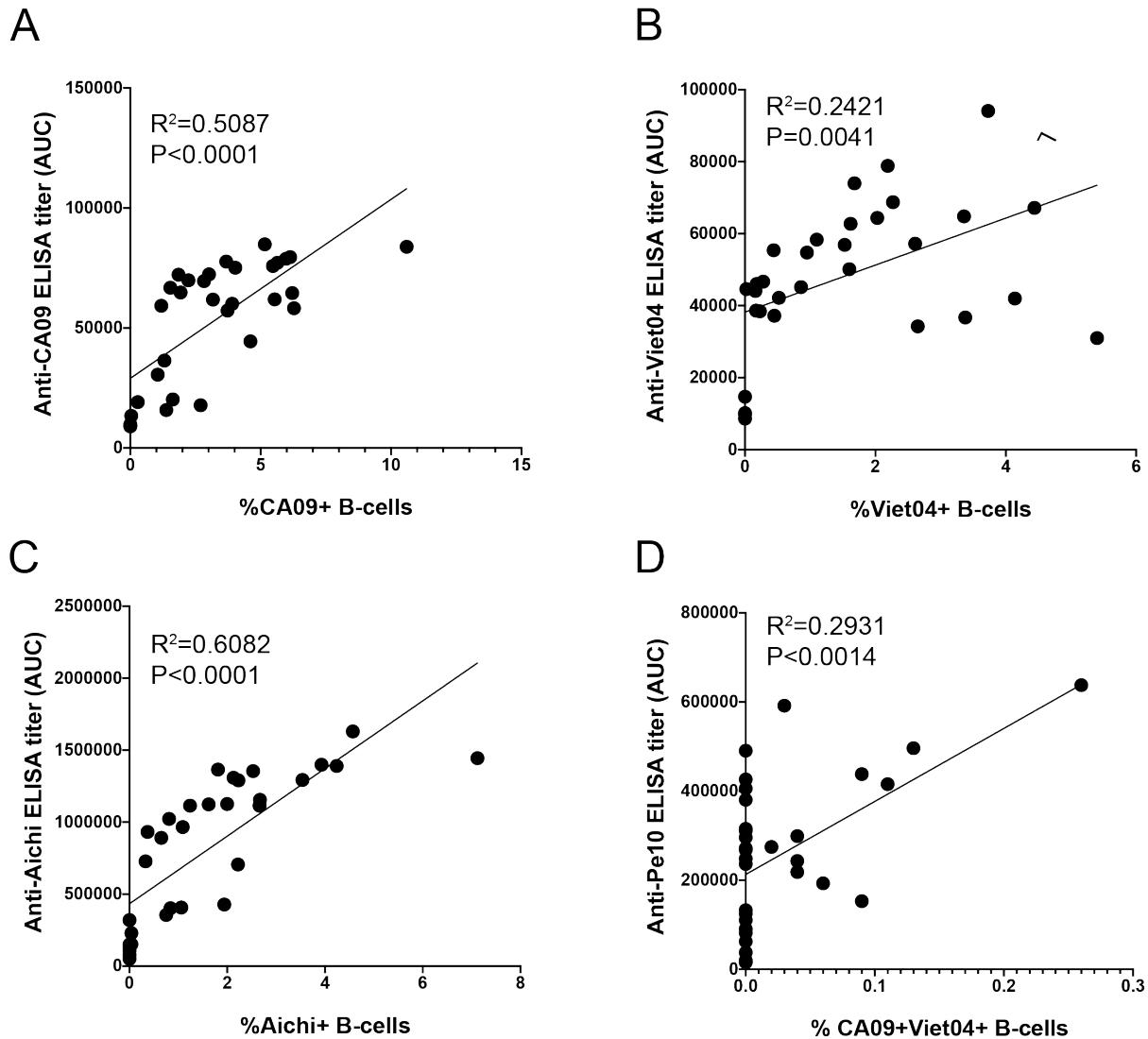


**Fig 8. B-cell responses induced by Mosaic VLP and NP immunization.** Flow Cytometry Analysis of IgG+ B-cells isolated from splenocytes as described in Fig 6A. **A.** Percent CA09+ (Y98F), Viet04+ (Y98F), Aichi+ (Y98F), and Sh13+ (Y98F) B-cells plotted for each group. **B.** Cross-reactive B-cell compartment: CA09+/Viet04+, CA09+/Aichi+, Sh13+/Aichi+ B-cell plotted for each group. Significant differences between groups represented by horizontal lines are indicated by asterisks:  $p<0.05=*$ ,  $p<0.01=**$ ,  $p<0.001=***$ , and  $p<0.0001=****$ . Differences with no significance are not shown, and significant differences between HA-NPs and mi3-NPs, HA-VLPs, and Ap205-VLPs, and HA-mi3s and HA-VLPs are also not shown.

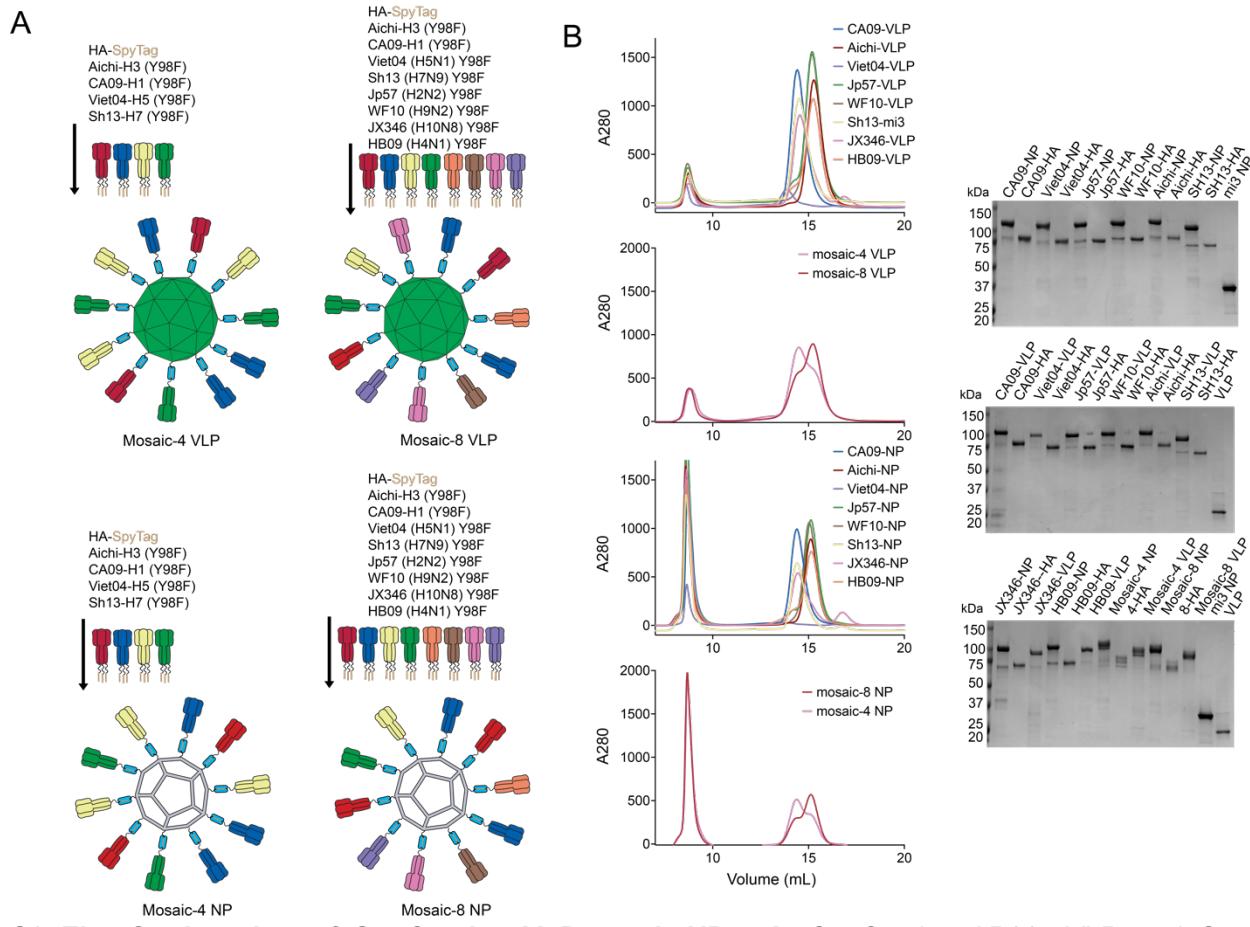
## Supporting Information



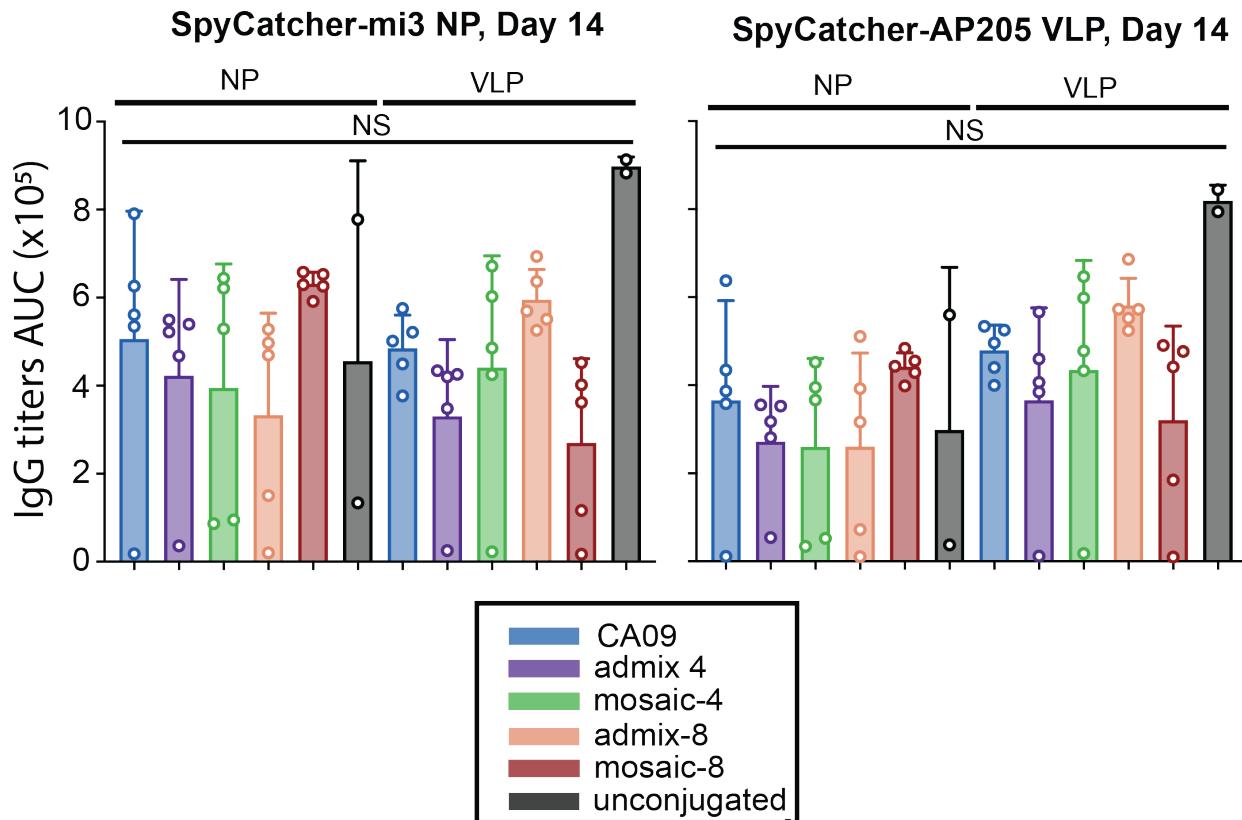
**S1 Fig. Correlation of ELISA AUC titers to live and pseudoviral neutralization titers.** Pearson correlation of Day45 serum ELISA AUC titers and viral neutralization titers for **A. CA09 H1N1** **B. Viet04 H5** **C. Aichi H3N2** **D. Sh13 H7** **E. JX346 H10**.



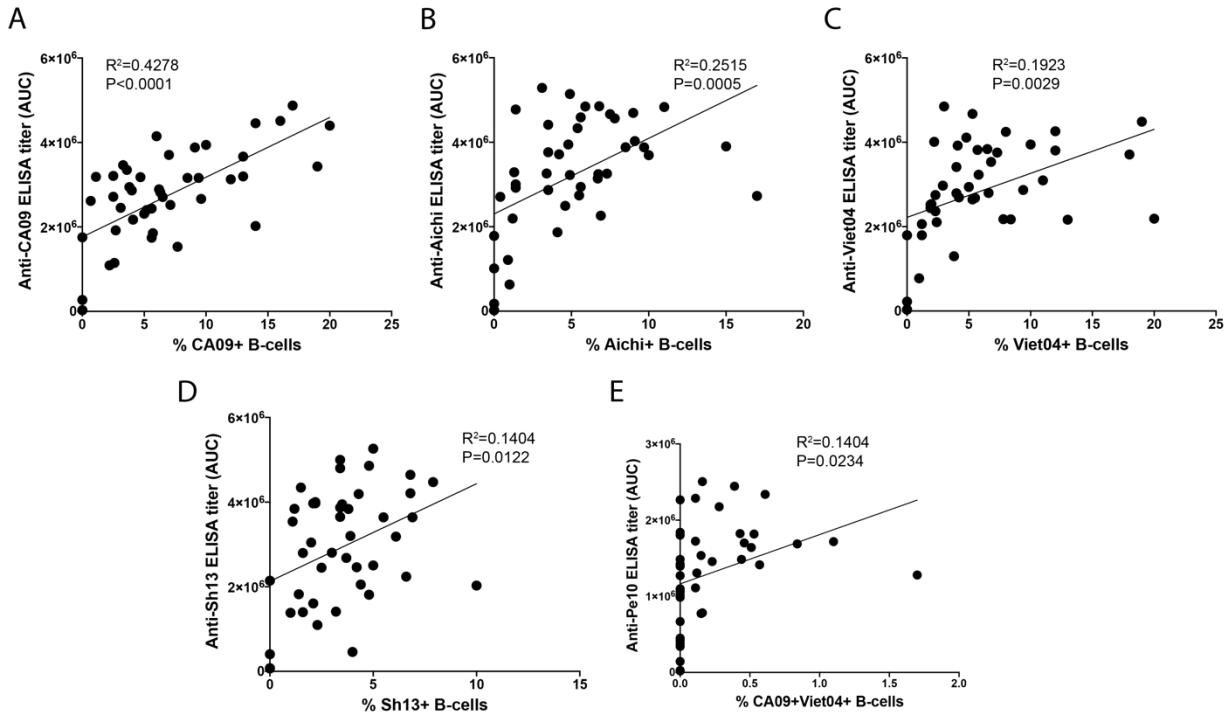
**S2 Fig. Correlation of ELISA AUC titers to antigen-specific B-cell populations.** **A.** Pearson correlation of Day45 CA09+ B-cell population to serum anti-CA09 ELISA AUC titers. **B.** Pearson correlation of Day45 Viet04+ B-cell population to serum anti-Viet04 ELISA AUC titers. **C.** Pearson correlation of Day45 Aichi+ B-cell population to serum anti-Aichi ELISA AUC titers. **D.** Pearson correlation of Day45 CA09+Viet04+ B-cell population to serum anti-Pe10 ELISA AUC titers.



**S3 Fig. Conjugation of SpyCatcher-VLPs and -NPs. A.** SpyCatcher-AP205-VLP and SpyCatcher-mi3 conjugations with SpyTagged-HA trimers. **B.** Purification of homotypic and mosaic SpyCatcher-VLPs and SpyCatcher-mi3s. Left: SEC separation of conjugated NPs from free HA trimers. Right: Reducing SDS-PAGE analysis of NPs and purified HAs.



**S4 Fig. Immunizations using Mosaic VLPs and NPs.** Serum antibody response to HA shown by ELISA binding as area under the curve (AUC) of Day 14 serum to SpyCatcher NP and VLP particles, with means and standard deviations represented by rectangles and horizontal lines, respectively. Homotypic strains that were present on the mosaic NPs and heterotypic strains that were not present are indicated by the blue and red rectangles, respectively, above the ELISA data.



**S5 Fig. Correlation of ELISA AUC titers to antigen-specific B-cell populations.** **A.** Pearson correlation of Day28 CA09+ B-cell population to serum anti-CA09 ELISA AUC titers. **B.** Pearson correlation of Day28 Aichi+ B-cell population to serum anti-Aichi ELISA AUC titers. **C.** Pearson correlation of Day28 Viet04+ B-cell population to serum anti-Viet04 ELISA AUC titers. **D** Pearson correlation of Day28 Sh13+ B-cell population to serum anti-Sh13 ELISA AUC titers. **E.** Pearson correlation of Day45 CA09+Viet04+ B-cell population to serum anti-Pe10 ELISA AUC titers.

Chapter 3: Mosaic nanoparticles elicit cross-reactive immune responses to zoonotic coronaviruses in mice

This chapter describes the development of mosaic nanoparticles that co-display RBDs from eight different sarbecovirus strains. These mosaic-RBD nanoparticles were designed in order to elicit antibody responses to conserved epitopes shared between the sarbecovirus RBDs. Immunizing with the mosaic-RBD nanoparticles elicited a polyclonal immune response that was cross-reactive and cross-neutralizing to all the sarbecovirus strains tested, including strains both present and not present on the particles. Furthermore, the antibody responses elicited by mosaic-RBD immunized mice were significantly more cross-reactive than immunization with a homotypic SARS-2 RBD nanoparticle. This study offers a viable strategy for a universal coronavirus vaccine.

My contribution to this work was as the lead researcher of the project. I conceived and designed the study, prepared the reagents, analyzed the data, and wrote the paper.

This work was published as:

**Cohen, A.A.**, Gnanapragasam, P.N.P., Lee, Y.E., Hoffman, P.R., Ou, S., Kakutani, L.M., Keeffe, J.R., Wu, H.-J., Howarth, M., West, A.P., et al. (2021). Mosaic nanoparticles elicit cross-reactive immune responses to zoonotic coronaviruses in mice. *Science* 371, 735–741.

[10.1126/science.abf6840](https://doi.org/10.1126/science.abf6840)

**Abstract:**

Protection against SARS-CoV-2 and SARS-related emergent zoonotic coronaviruses is urgently needed. We made homotypic nanoparticles displaying the receptor-binding domain (RBD) of SARS-CoV-2 or co-displaying SARS-CoV-2 RBD along with RBDs from animal betacoronaviruses that represent threats to humans (mosaic nanoparticles; 4-8 distinct RBDs). Mice immunized with RBD-nanoparticles, but not soluble antigen, elicited cross-reactive binding and neutralization responses. Mosaic-RBD-nanoparticles elicited antibodies with superior cross-reactive recognition of heterologous RBDs compared to sera from immunizations with homotypic SARS-CoV-2–RBD-nanoparticles or COVID-19 convalescent human plasmas. Moreover, sera from mosaic-RBD–immunized mice neutralized heterologous pseudotyped coronaviruses equivalently or better after priming than sera from homotypic SARS-CoV-2–RBD-nanoparticle immunizations, demonstrating no immunogenicity loss against particular RBDs resulting from co-display. A single immunization with mosaic-RBD-nanoparticles provides a potential strategy to simultaneously protect against SARS-CoV-2 and emerging zoonotic coronaviruses.

Main Text:

SARS-CoV-2, a newly-emergent betacoronavirus, resulted in a global pandemic in 2020, infecting millions and causing the respiratory disease COVID-19 (1, 2). Two other zoonotic betacoronaviruses, SARS-CoV and MERS-CoV, also resulted in outbreaks within the last 20 years (3). All three viruses presumably originated in bats (4), with SARS-CoV and MERS-CoV adapting to intermediary animal hosts before jumping to humans. SARS-like viruses circulate in bats and serological surveillance of people living near caves where bats carry diverse coronaviruses demonstrates direct transmission of SARS-like viruses with pandemic potential (5), suggesting a pan-coronavirus vaccine is needed to protect against future outbreaks and pandemics. In particular, the bat WIV1 and SCH014 strains are thought to represent an ongoing threat to humans (6, 7).

Most current SARS-CoV-2 vaccine candidates include the spike trimer (S), the viral protein that mediates target cell entry after one or more of its receptor-binding domains (RBDs) adopt an “up” position to bind a host receptor (Fig. 1A). The RBDs of human coronaviruses SARS-CoV-2, SARS-CoV, HCoV-NL63, and related animal coronaviruses (WIV1 and SCH014) use angiotensin-converting enzyme 2 (ACE2) as their host receptor (1, 8, 9), while other coronaviruses use receptors such as dipeptidyl peptidase 4 (10) or sialic acids (11, 12). Consistent with its function in viral entry, S is the primary target of neutralizing antibodies (13-22), with many targeting the RBD (14-18, 21-26).

Multivalent display of antigen enhances B-cell responses and can provide longer-lasting immunity than monovalent antigens (27, 28), thus protein-based vaccine candidates often involve a nanoparticle that enables antigen multimerization. Many nanoparticles and coupling strategies have been explored for vaccine design (29), with “plug and display” strategies being especially useful (30, 31).

In one such approach, multiple copies of an engineered protein domain called SpyCatcher fused to subunits of a virus-like particle form spontaneous isopeptide bonds to purified antigens tagged with a 13-residue SpyTag (29-32). The SpyCatcher-SpyTag system was used to prepare multimerized SARS-CoV-2 RBD or S trimer that elicited high titers of neutralizing antibodies (33, 34). Although promising for protection against SARS-CoV-2, coronavirus reservoirs in bats suggest future cross-species transmission (6, 7, 35), necessitating a vaccine that protects against emerging coronaviruses as well as SARS-CoV-2. Here we prepared SpyCatcher003-mi3 nanoparticles (31, 36) simultaneously displaying SpyTagged RBDs from human and animal coronaviruses to evaluate whether mosaic particles can elicit cross-reactive antibody responses, as previously demonstrated for influenza head domain mosaic particles (37). We show that mice immunized with homotypic or mosaic nanoparticles produced broad binding and neutralizing responses, in contrast to plasma antibodies elicited in humans by SARS-CoV-2 infection. Moreover, mosaic nanoparticles showed enhanced heterologous binding and neutralization properties against human and bat SARS-like betacoronaviruses (sarbecoviruses) compared with homotypic SARS-CoV-2 nanoparticles.

We used a study of sarbecovirus RBD receptor usage and cell tropism (38) to guide our choice of RBDs for co-display on mosaic particles. From 29 RBDs that were classified into distinct clades (clades 1, 2, 1/2, and 3) (38), we identified diverse RBDs from SARS-CoV, WIV1, and SHC014 (clade 1), SARS-CoV-2 (clade 1/2), Rs4081, Yunnan 2011 (Yun11), and Rf1 (clade 2), and BM48-31 (clade 3), of which SARS-CoV-2 and SARS-CoV are human coronaviruses and the rest are bat viruses originating in China or Bulgaria (BM48-31). We also included RBDs from the GX pangolin clade 1/2 coronavirus (referred to here as pang17) (39), RaTG13, the bat clade 1/2 virus most closely related to SARS-CoV-2 (40), RmYN02, a clade 2 bat virus from China (41), and BtKY72, a

Kenyan bat clade 3 virus (42). Mapping of the sequence conservation across selected RBDs showed varying degrees of sequence identity (68-95%), with highest sequence variability in residues corresponding to the SARS-CoV-2 ACE2 receptor-binding motif (Fig. 1A-D; fig. S1). We chose 8 of the 12 RBDs for making three types of mosaic nanoparticles: mosaic-4a (coupled to SARS-2, RaTG13, SHC014, and Rs4081 RBDs), mosaic-4b (coupled to pang17, RmYN02, RF1, and WIV1 RBDs), and mosaic-8 (coupled to all eight RBDs), and compared them with homotypic mi3 particles constructed from SARS-CoV-2 RBD alone (homotypic SARS-2). RBDs from SARS, Yun11, BM-4831, and BtKY72, which were not coupled to mosaic particles, were used to evaluate sera for cross-reactive responses.

SpyTag003-RBDs were coupled to SpyCatcher003-mi3 (60 potential conjugation sites) (36, 43) to make homotypic and mosaic nanoparticles (Fig 2A). Particles were purified by size exclusion chromatography (SEC) and analyzed by SDS-PAGE, revealing monodisperse SEC profiles and nearly 100% conjugation (Fig. 2B,C). Representative RBDs were conjugated to SpyCatcher003-mi3 with similar or identical efficiencies (fig. S2), suggesting that mosaic particles contained approximately equimolar mixtures of different RBDs.

We immunized mice with either soluble SARS-CoV-2 spike trimer (SARS-2 S), nanoparticles displaying only SARS-2 RBD (homotypic SARS-2), nanoparticles co-displaying RBDs (mosaic-4a, mosaic-4b, mosaic-8), or unconjugated nanoparticles (mi3). IgG responses were evaluated after prime or boost immunizations (Fig. 3A) by ELISA against SARS-2 S (Fig. 3B) or a panel of RBDs (Fig. 3C-F; fig. S3). Sera from unconjugated nanoparticle-immunized animals (black in Fig. 3, fig. S3)

showed no responses above background. Anti-SARS-2 S trimer and anti-SARS-2 RBD serum responses were similar (Fig. 3B,C), demonstrating that antibodies elicited against RBDs can access their epitopes on SARS-2 S trimer. We also conducted in vitro neutralization assays using a pseudotyped virus assay that quantitatively correlates with authentic virus neutralization (44) for strains known to infect 293T<sub>ACE2</sub> target cells (SARS-CoV-2, SARS, WIV1 and SHC104). Neutralization and ELISA titers were significantly correlated (fig. S4), thus suggesting ELISAs are predictive of neutralization results when pseudotyped neutralization assays were not possible due to unknown viral entry receptor usage.

Mice immunized with soluble SARS-2 S trimer (brown bars) showed no binding or neutralization except for autologous responses against SARS-2 after boosting (Fig. 3C-F). By contrast, sera from RBD-nanoparticle-immunized animals (red, green, yellow, and blue bars) exhibited binding to all RBDs (Fig. 3C-F; fig. S3A) and neutralization against all four strains after boosting (Fig. 3C-E), consistent with increased immunogenicities of multimerized antigen on nanoparticles versus soluble antigen (27, 28). Homotypic SARS-2 nanoparticles, but not soluble SARS-2 trimer, induced heterologous responses to zoonotic RBDs and neutralization of heterologous coronaviruses (Fig. 3D-F). To address whether co-display of SARS-2 RBD along with other RBDs on mosaic-4a and mosaic-8 versus homotypic display of SARS-2 RBD (homotypic SARS-2) diminished anti-SARS-2 responses, we compared SARS-2-specific ELISA and neutralization titers for mosaic versus homotypic immunizations (Fig. 3C): there were no significant differences in IgG anti-SARS-2 titers for animals immunized with homotypic (red in Fig. 3C) versus mosaic nanoparticles (green and blue in Fig. 3C). Thus there was no advantage of immunization with a homotypic RBD-nanoparticle versus

a mosaic-nanoparticle that included SARS-2 RBD in terms of the magnitude of immune responses against SARS-2.

We next compared serum responses against matched RBDs (RBDs present on an injected nanoparticle; gray horizontal shading) versus mismatched RBDs (RBDs not present on injected nanoparticle; red horizontal shading) (Fig. 3; fig. S3). Although SARS-2 RBD was not presented on mosaic-4b, antibody titers elicited by mosaic-4b immunization (yellow) were not significantly different than titers elicited by matched nanoparticle immunizations (homotypic SARS-2 (red), mosaic-4a (green), and mosaic-8 (blue)), and sera from boosted mosaic-4b-immunized mice neutralized SARS-2 pseudovirus (Fig. 3C). In other matched versus mismatched comparisons, sera showed binding and neutralization of SHC014 and WIV1 regardless of whether these RBDs were included on the injected nanoparticle (Fig. 3D), underscoring sharing of common epitopes among RBDs (Fig. 1A).

Demonstrating advantages of mosaic versus homotypic SARS-2 nanoparticles, sera from mosaic-8-immunized mice bound SHC014 and WIV1 RBDs significantly better after the prime than sera from homotypic SARS-2-immunized mice and retained better binding to SHC014 RBD after boosting (Fig. 3D). Thus the potential increased avidity of the homotypic SARS-2 nanoparticle displaying only one type of RBD over the mosaic-8 nanoparticles did not confer increased breadth. Moreover, mosaic-8-immunized and boosted sera were 7-44-fold more potent than sera from homotypic SARS-2-immunized animals in neutralizing SHC014 and WIV1 (Fig. 3D). Neutralization of the SHC014 and WIV1 pseudoviruses by mosaic-8 sera suggests that combining RBDs on a mosaic

nanoparticle does not diminish the immune response against a particular RBD, also suggested by ELISA binding of sera to Rs4081 and RaTG13 (fig. S3A,B).

To further address whether RBD-nanoparticles elicited antibodies that recognized totally mismatched strains and SARS-CoV-2 RBD mutants, we evaluated sera for binding to SARS, Yun11, BM-4831, and BtKY72 RBDs (Fig. 3E,F), SARS-2 RBD mutants (fig. S3C), MERS-CoV RBD (fig. S3D), and for neutralization in SARS pseudovirus assays (Fig. 3E). We found no reductions in SARS-2 RBD binding as a result of mutations (Y453F, the “Danish mink variant” (45) or a Q493K/Q498Y/P499T triple mutant (46)) (fig. S3C), no binding of any elicited sera to MERS-CoV RBD (fig. S3D), and higher and more cross-reactive antibody responses for mosaic immunizations compared with homotypic SARS-2 immunizations: e.g., mosaic-8-primed and boosted animals showed significantly higher titers against SARS RBD than sera from homotypic SARS-2-immunized mice (Fig. 3E). After the prime, sera from the homotypic SARS-2-immunized animals did not neutralize SARS, whereas the mosaic-4b and mosaic-8 sera were neutralizing (Fig. 3E), perhaps facilitated by these nanoparticles including WIV1 RBD, which is related by 95% amino acid identity to SARS RBD (Fig. 1D). After boosting, SARS-2 and mosaic-4a sera were also neutralizing, although titers were ~4-fold lower than for mosaic-8-immunized animals (Fig. 3E). ELISA titers against other mismatched RBDs (Yun11, BM-4831, BtKY72) were significantly higher for sera collected after mosaic-8 priming compared to sera from homotypic SARS-2 priming, and heightened binding was retained after boosting (Fig. 3F). Thus mosaic nanoparticles, particularly mosaic-8, induce higher antibody titers against mismatched RBDs than homotypic SARS-2 nanoparticles, again favoring the co-display approach for inducing broader anti-coronavirus responses, especially after a single prime.

We investigated the potential for cross-reactive recognition using flow cytometry to ask whether B-cell receptors on IgG+ splenic B-cells from RBD-nanoparticle-boosted animals could simultaneously recognize RBDs from SARS-2 and Rs4081 (related by 70% sequence identity) (Fig. 1D; fig. S5). Whereas control animals were negative, all other groups showed B-cells that recognized SARS-2 and Rs4081 RBDs simultaneously, suggesting the existence of antibodies that cross-react with both RBDs (fig. S5E).

To compare antibodies elicited by RBD-nanoparticle immunization to antibodies elicited by SARS-CoV-2 infection, we repeated ELISAs against the RBD panel using IgGs from COVID-19 plasma donors (47) (Fig. 4). Most of the convalescent plasmas showed detectable binding to SARS-2 RBD (Fig. 4A). However, binding to other sarbecovirus RBDs (RaTG13, SHC014, WIV1, Rs4081 and BM-4831) was significantly weaker than binding to SARS 2 RBD, with many human plasma IgGs showing no binding above background (Fig. 4B-G). In addition, although convalescent plasma IgGs neutralized SARS-CoV-2 pseudoviruses, they showed weak or no neutralization of SARS, SHC014, or WIV1 pseudoviruses (Fig. 4H). These results are consistent with little to no cross-reactive recognition of RBDs from zoonotic coronavirus strains resulting from SARS-CoV-2 infection in humans.

In conclusion, we confirmed that multimerization of RBDs on nanoparticles enhances immunogenicity compared with soluble antigen (33, 48) and further showed that homotypic SARS-2 nanoparticle immunization produced IgG responses that bound zoonotic RBDs and neutralized heterologous coronaviruses after boosting. By contrast, soluble SARS-2 S immunization and natural infection with SARS-CoV-2 resulted in weak or no heterologous responses in plasmas. Co-display of SARS-

2 RBD along with diverse RBDs on mosaic nanoparticles showed no disadvantages for eliciting neutralizing antibodies against SARS-CoV-2 compared with homotypic SARS-2 nanoparticles, suggesting mosaic nanoparticles as a candidate vaccine to protect against COVID-19. Furthermore, compared with homotypic SARS-2 RBD particles, the mosaic co-display strategy demonstrated advantages for eliciting neutralizing antibodies against zoonotic sarbecoviruses, thus potentially also providing protection against emerging coronaviruses with human spillover potential. Neutralization of matched and mismatched strains was observed after mosaic priming, suggesting a single injection of a mosaic-RBD nanoparticle might be sufficient in a vaccine. Since COVID-19 convalescent plasmas showed little to no recognition of coronavirus RBDs other than SARS-CoV-2, COVID-19-induced immunity in humans may not protect against another emergent coronavirus. However, the mosaic nanoparticles described here could be used as described or easily adapted to present RBDs from newly-discovered zoonotic coronaviruses.

## References and Notes

1. P. Zhou *et al.*, A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 579, 270-273 (2020).
2. F. Wu *et al.*, A new coronavirus associated with human respiratory disease in China. *Nature* 579, 265-269 (2020).
3. E. de Wit, N. van Doremalen, D. Falzarano, V. J. Munster, SARS and MERS: recent insights into emerging coronaviruses. *Nat Rev Microbiol* 14, 523-534 (2016).
4. W. Li *et al.*, Bats are natural reservoirs of SARS-like coronaviruses. *Science* 310, 676-679 (2005).
5. N. Wang *et al.*, Serological Evidence of Bat SARS-Related Coronavirus Infection in Humans, China. *Virologica Sinica* 33, 104-107 (2018).
6. V. D. Menachery *et al.*, A SARS-like cluster of circulating bat coronaviruses shows potential for human emergence. *Nature Medicine* 21, 1508-1513 (2015).
7. V. D. Menachery *et al.*, SARS-like WIV1-CoV poised for human emergence. *Proceedings of the National Academy of Sciences* 113, 3048-3053 (2016).
8. W. Li *et al.*, Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* 426, 450-454 (2003).

9. M. Hoffmann *et al.*, SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* 181, 271-280 e278 (2020).
10. V. S. Raj *et al.*, Dipeptidyl peptidase 4 is a functional receptor for the emerging human coronavirus-EMC. *Nature* 495, 251-254 (2013).
11. R. Vlasak, W. Luytjes, W. Spaan, P. Palese, Human and bovine coronaviruses recognize sialic acid-containing receptors similar to those of influenza C viruses. *Proceedings of the National Academy of Sciences* 85, 4526-4529 (1988).
12. X. Huang *et al.*, Human Coronavirus HKU1 Spike Protein Uses O-Acetylated Sialic Acid as an Attachment Receptor Determinant and Employs Hemagglutinin-Esterase Protein as a Receptor-Destroying Enzyme. *Journal of Virology* 89, 7202-7213 (2015).
13. T. S. Fung, D. X. Liu, Human Coronavirus: Host-Pathogen Interaction. *Annu Rev Microbiol* 73, 529-557 (2019).
14. P. J. M. Brouwer *et al.*, Potent neutralizing antibodies from COVID-19 patients define multiple targets of vulnerability. *Science* 369, 643-650 (2020).
15. Y. Cao *et al.*, Potent neutralizing antibodies against SARS-CoV-2 identified by high-throughput single-cell sequencing of convalescent patients' B cells. *Cell*, (2020).
16. C. Kreer *et al.*, Longitudinal Isolation of Potent Near-Germline SARS-CoV-2-Neutralizing Antibodies from COVID-19 Patients. *Cell*, (2020).
17. L. Liu *et al.*, Potent neutralizing antibodies against multiple epitopes on SARS-CoV-2 spike. *Nature*, (2020).
18. D. F. Robbiani *et al.*, Convergent antibody responses to SARS-CoV-2 in convalescent individuals. *Nature* 584, 437-442 (2020).
19. R. Shi *et al.*, A human neutralizing antibody targets the receptor-binding site of SARS-CoV-2. *Nature* 584, 120-124 (2020).
20. S. J. Zost *et al.*, Rapid isolation and profiling of a diverse panel of human monoclonal antibodies targeting the SARS-CoV-2 spike protein. *Nat Med*, (2020).
21. T. F. Rogers *et al.*, Rapid isolation of potent SARS-CoV-2 neutralizing antibodies and protection in a small animal model. *Science*, (2020).
22. E. Seydoux *et al.*, Analysis of a SARS-CoV-2-Infected Individual Reveals Development of Potent Neutralizing Antibodies with Limited Somatic Mutation. *Immunity* 53, 98-105 e105 (2020).
23. S. J. Zost *et al.*, Potently neutralizing and protective human antibodies against SARS-CoV-2. *Nature* 584, 443-449 (2020).
24. C. O. Barnes *et al.*, SARS-CoV-2 neutralizing antibody structures inform therapeutic strategies. *Nature*, (2020).
25. D. Pinto *et al.*, Cross-neutralization of SARS-CoV-2 by a human monoclonal SARS-CoV antibody. *Nature* 583, 290-295 (2020).
26. 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*, (2020).
27. J. López-Sagasta, E. Malito, R. Rappuoli, M. J. Bottomley, Self-assembling protein nanoparticles in the design of vaccines. *Computational and Structural Biotechnology Journal* 14, 58-68 (2016).

28. M. K. Slifka, I. J. Amanna, Role of Multivalency and Antigenic Threshold in Generating Protective Antibody Responses. *Frontiers in Immunology* 10, (2019).
29. K. D. Brune, M. Howarth, New Routes and Opportunities for Modular Construction of Particulate Vaccines: Stick, Click, and Glue. *Front Immunol* 9, 1432 (2018).
30. K. D. Brune *et al.*, Plug-and-Display: decoration of Virus-Like Particles via isopeptide bonds for modular immunization. *Scientific reports* 6, 19234 (2016).
31. T. U. J. Bruun, A. C. Andersson, S. J. Draper, M. Howarth, Engineering a Rugged Nanoscaffold To Enhance Plug-and-Display Vaccination. *ACS Nano* 12, 8855-8866 (2018).
32. B. Zakeri *et al.*, Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin. *Proc Natl Acad Sci U S A* 109, E690-697 (2012).
33. T. K. Tan *et al.*, A COVID-19 vaccine candidate 1 using SpyCatcher multimerization of the SARS-CoV-2 spike protein receptor-binding domain induces potent neutralising antibody responses. *bioRxiv*, (2020).
34. B. Zhang *et al.*, A platform incorporating trimeric antigens into self-assembling nanoparticles reveals SARS-CoV-2-spike nanoparticles to elicit substantially higher neutralizing responses than spike alone. *Scientific reports* 10, (2020).
35. C. Drosten *et al.*, Discovery of a rich gene pool of bat SARS-related coronaviruses provides new insights into the origin of SARS coronavirus. *PLOS Pathogens* 13, e1006698 (2017).
36. R. Rahikainen *et al.*, Overcoming Symmetry Mismatch in Vaccine Nanoassembly through Spontaneous Amidation. *Angewandte Chemie International Edition*, (2020).
37. M. Kanekiyo *et al.*, Mosaic nanoparticle display of diverse influenza virus hemagglutinins elicits broad B cell responses. *Nat Immunol* 20, 362-372 (2019).
38. M. Letko, A. Marzi, V. Munster, Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. *Nature Microbiology* 5, 562-569 (2020).
39. T. T.-Y. Lam *et al.*, Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins. *Nature* 583, 282-285 (2020).
40. K. G. Andersen, A. Rambaut, W. I. Lipkin, E. C. Holmes, R. F. Garry, The proximal origin of SARS-CoV-2. *Nature Medicine* 26, 450-452 (2020).
41. H. Zhou *et al.*, A Novel Bat Coronavirus Closely Related to SARS-CoV-2 Contains Natural Insertions at the S1/S2 Cleavage Site of the Spike Protein. *Current Biology* 30, 2196-2203.e2193 (2020).
42. Y. Tao, S. Tong, K. M. Stedman, Complete Genome Sequence of a Severe Acute Respiratory Syndrome-Related Coronavirus from Kenyan Bats. *Microbiology Resource Announcements* 8, (2019).
43. A. H. Keeble *et al.*, Approaching infinite affinity through engineering of peptide–protein interaction. *Proceedings of the National Academy of Sciences* 116, 26523-26533 (2019).
44. F. Schmidt *et al.*, Measuring SARS-CoV-2 neutralizing antibody activity using pseudotyped and chimeric viruses. *Journal of Experimental Medicine* 217, (2020).
45. L. van Dorp *et al.*, Recurrent mutations in SARS-CoV-2 genomes isolated from mink point to rapid host-adaptation. *bioRxiv*, (2020).

46. S. R. Leist *et al.*, A Mouse-Adapted SARS-CoV-2 Induces Acute Lung Injury and Mortality in Standard Laboratory Mice. *Cell* 183, 1070-1085.e1012 (2020).
47. C. O. Barnes *et al.*, Structures of Human Antibodies Bound to SARS-CoV-2 Spike Reveal Common Epitopes and Recurrent Features of Antibodies. *Cell* 182, 828-842 e816 (2020).
48. A. C. Walls *et al.*, Elicitation of Potent Neutralizing Antibody Responses by Designed Protein Nanoparticle Vaccines for SARS-CoV-2. *Cell* 183, 1367-1382.e1317 (2020).
49. M. Landau *et al.*, ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res* 33, W299-302 (2005).
50. H. B. Gristick *et al.*, Natively glycosylated HIV-1 Env structure reveals new mode for antibody recognition of the CD4-binding site. *Nat Struct Mol Biol* 23, 906-915 (2016).
51. F. Sievers *et al.*, Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 7, 539 (2011).
52. S. Guindon *et al.*, New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 59, 307-321 (2010).
53. C.-L. Hsieh *et al.*, Structure-based design of prefusion-stabilized SARS-CoV-2 spikes. *Science* 369, 1501-1505 (2020).
54. A. A. Cohen *et al.*, Construction, characterization, and immunization of nanoparticles that display a diverse array of influenza HA trimers. *bioRxiv*, (2020).
55. D. Angeletti *et al.*, Defining B cell immunodominance to viruses. *Nature Immunology* 18, 456-463 (2017).
56. K. H. D. Crawford *et al.*, Protocol and Reagents for Pseudotyping Lentiviral Particles with SARS-CoV-2 Spike Protein for Neutralization Assays. *Viruses* 12, 513 (2020).
57. A. P. West, Jr. *et al.*, Computational analysis of anti-HIV-1 antibody neutralization panel data to identify potential functional epitope residues. *Proc Natl Acad Sci U S A* 110, 10598-10603 (2013).
58. T. N. Starr *et al.*, Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding. *Cell* 182, 1295-1310.e1220 (2020).

## Acknowledgements

We thank Karl Brune (Genie Biotech) for advice about mi3 production, Jesse Bloom (Fred Hutchinson) and Paul Bieniasz (Rockefeller University) for neutralization assay reagents, Jost

Vielmetter and Caltech's Beckman Institute Protein Expression Center for protein production, Andrew Flyak for help with flow cytometry, Marta Murphy for figures, COVID-19 plasma donors and Drs. Barry Coller and Sarah Schlesinger and the Rockefeller University Hospital Clinical Research Support Office and nursing staff, and Andrew Flyak and Andrew DeLaitsch for critical reading of the manuscript.

**Funding:** This work was supported by NIH grant P01-AI138938-S1 (P.J.B. and M.C.N.), the Caltech Merkin Institute for Translational Research (P.J.B.), a George Mason University Fast Grant (P.J.B.), and the Medical Research Council (MR/P001351/1) (M.H.) (this UK-funded award is part of the EDCTP2 programme supported by the European Union). M.C.N. is a Howard Hughes Medical Institute Investigator.

**Author contributions:** A.A.C., C.O.B., and P.J.B. conceived and designed experiments. A.A.C., P.N.P.G., Y.E.L., P.R.H., S.O., and L.M.K. performed experiments, H-J.W. generated and validated SpyCatcher003-mi3, M.H. supervised the generation and validation of SpyCatcher003-mi3, A.A.C., J.R.K., A.P.W., C.O.B., M.C.N., and P.J.B. analyzed data and wrote the paper with contributions from other authors.

**Competing interests:** M.H. is an inventor on a patent on SpyTag/SpyCatcher (EP2534484) and a patent application on SpyTag003:SpyCatcher003 (UK Intellectual Property Office 1706430.4), as well as a SpyBiotech cofounder, shareholder, and consultant.

**Data and materials availability:** All data are available in the main text of the supplementary materials. Materials are available upon request to [bjorkman@caltech.edu](mailto:bjorkman@caltech.edu) with a signed Material Transfer Agreement.

**Supplementary content:** Materials and Methods, Figs. S1 to S5, References (51-58).

Figure 1

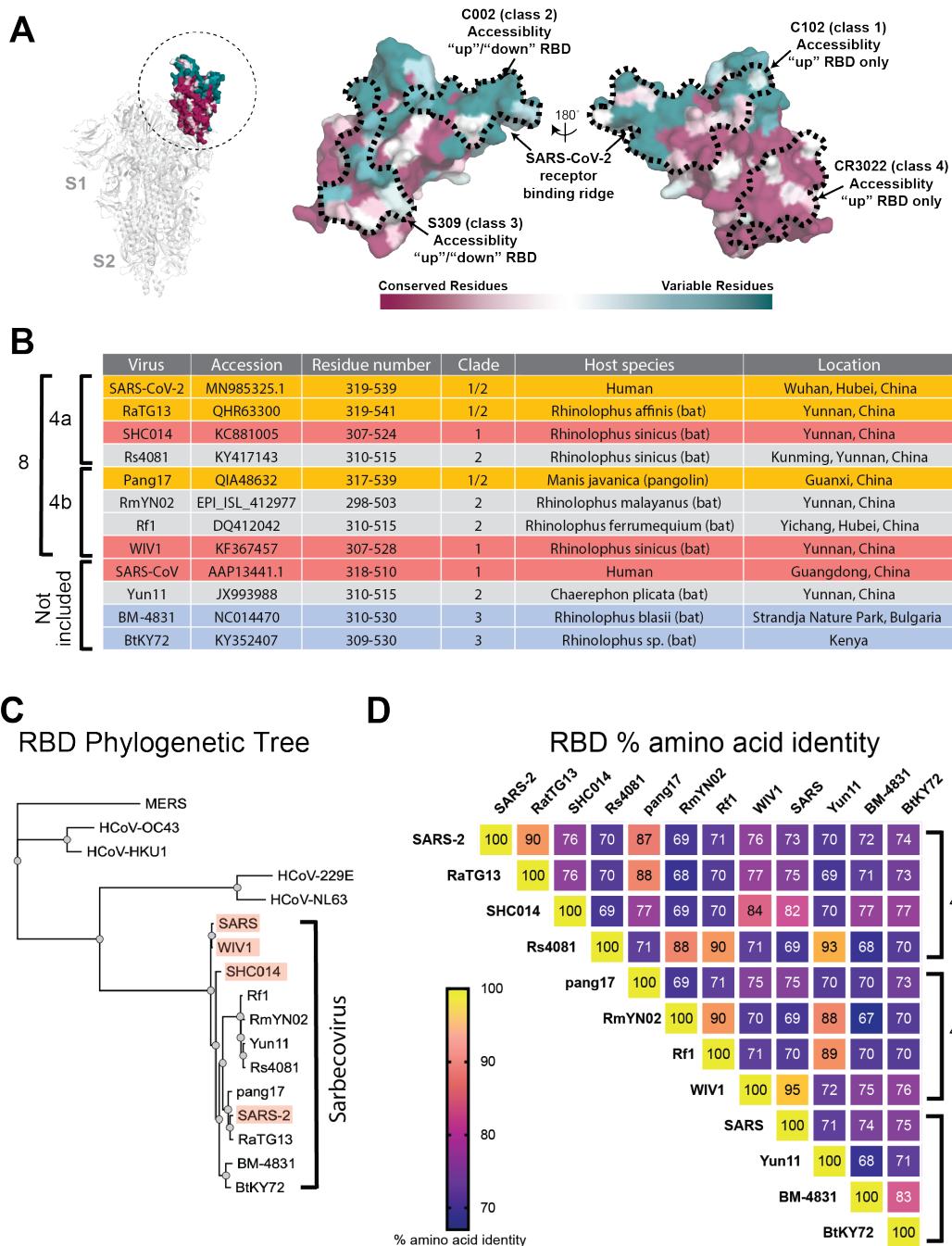


Figure 1. Properties of RBDs chosen for this study. (A) Left: Structure of SARS-CoV-2 S trimer (PDB 6VXX) with one RBD (dashed circle) in an “up” position. Middle and right: Sequence conservation of 12 RBDs calculated by the ConSurf Database (49) plotted on a surface representation of the RBD structure (PDB 7BZ5). Epitopes for representatives from defined classes of RBD-binding antibodies (class 1-class 4) (24) indicated by dashed lines. (B) Summary of properties of the viral strains from which the 12 sarbecovirus RBDs were derived. (C) Phylogenetic tree of human and selected other coronaviruses based on RBD protein sequences. Red shading indicates strains known to use ACE2 as a receptor. (D) Heat map showing percent amino acid sequence identities between 12 sarbecovirus RBDs.

Figure 2

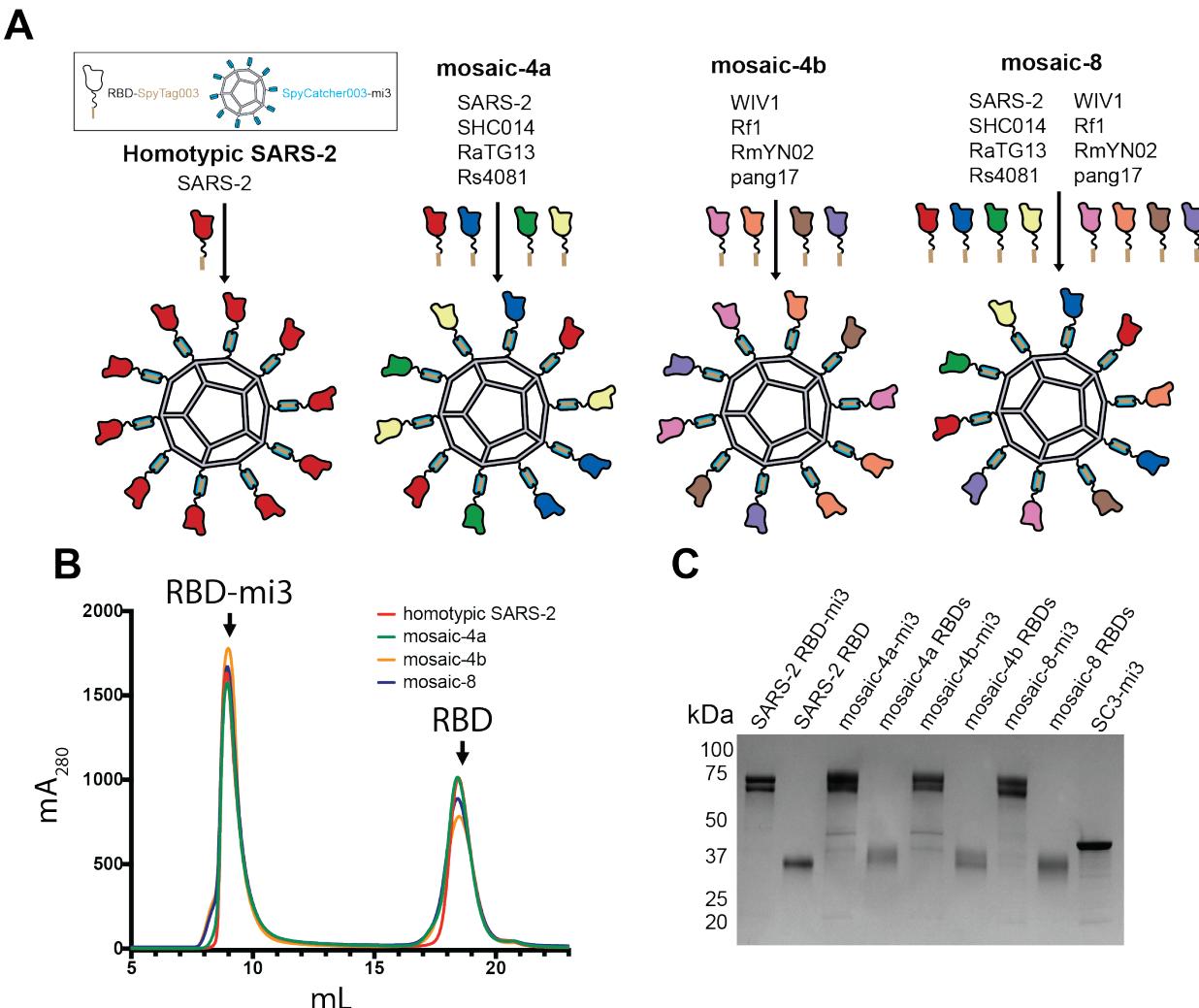


Figure 2. Construction of RBD nanoparticles. (A) Left: SpyTagged RBDs were attached to SpyCatcher003-mi3 to make a homotypic particle and three mosaic particles. 10 of 60 potential coupling sites on mi3 are shown for clarity. (B) SEC profile showing separation of RBD nanoparticles and free RBD proteins. (C) Coomassie-stained SDS-PAGE of RBD-coupled nanoparticles, free RBD proteins, and uncoupled SpyCatcher003-mi3 particles (SC3-mi3).



Figure 3

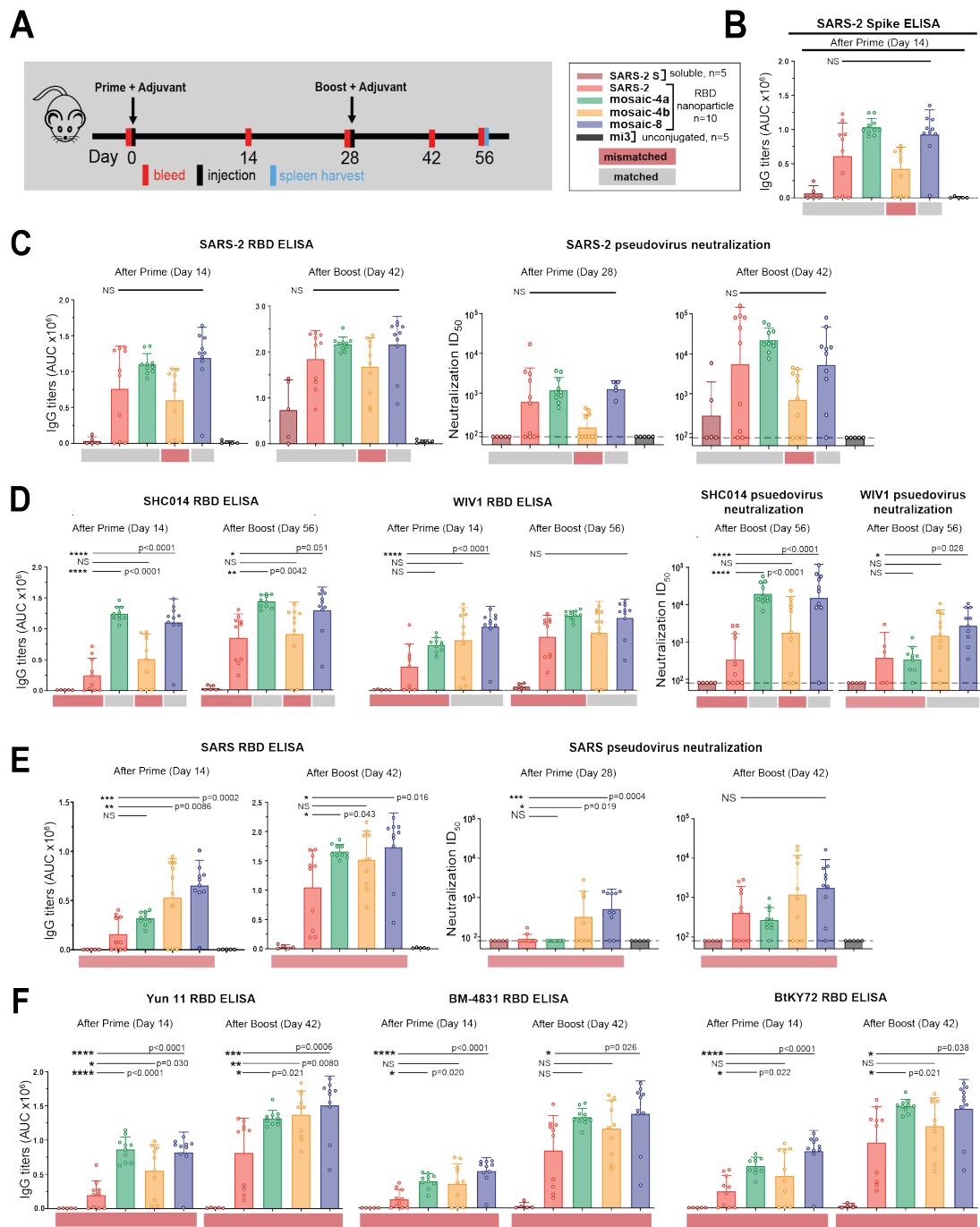


Figure 3. RBD nanoparticles induce cross-reactive IgG responses in immunized mice. Red and gray rectangles below ELISA and neutralization data represent mismatched strains (red; RBD from that strain was not present on the immunized particle) or matched strains (gray; RBD was present on the immunized particle). (A) Left: Immunization schedule. Adjuvant=AddaVax (Invivogen). Right: Key for immunizations; number of mice in each cohort is indicated. (B-F) Mice were immunized with soluble SARS-CoV-2 S trimer (SARS-2 S; brown bars), or the following nanoparticles: homotypic SARS-2 (red), mosaic-4a (green), mosaic-4b (yellow), mosaic-8 (blue), or unconjugated Spy-Catcher003-mi3 (mi3; black). ELISA data from serum IgG responses to SARS-2 spike trimer (B) or RBDs (C-F) shown as area under the curve (AUC). For C-E, neutralization potencies are presented as half-maximal inhibitory dilutions ( $ID_{50}$  values) of sera against the pseudoviruses from the indicated coronavirus strains. Dashed horizontal lines correspond to the lowest dilution representing the limit of detection. Each dot represents serum from one animal, with means and standard deviations for vaccinated cohorts represented by rectangles (mean) and horizontal lines (SD). Significant differences between groups linked by horizontal lines are indicated by asterisks and p-values. NS=not significant. (B-F) Neutralization and/or binding data for serum IgGs for recognition of (B) SARS-2 spike trimer, (C) SARS-2 RBD and SARS-2 pseudovirus, (D) SHC014 and WIV1 RBDs and corresponding pseudoviruses, (E) SARS RBD and SARS pseudovirus, (F) Yun 11, BM-4831, and BtKY72 RBDs.



Figure 4

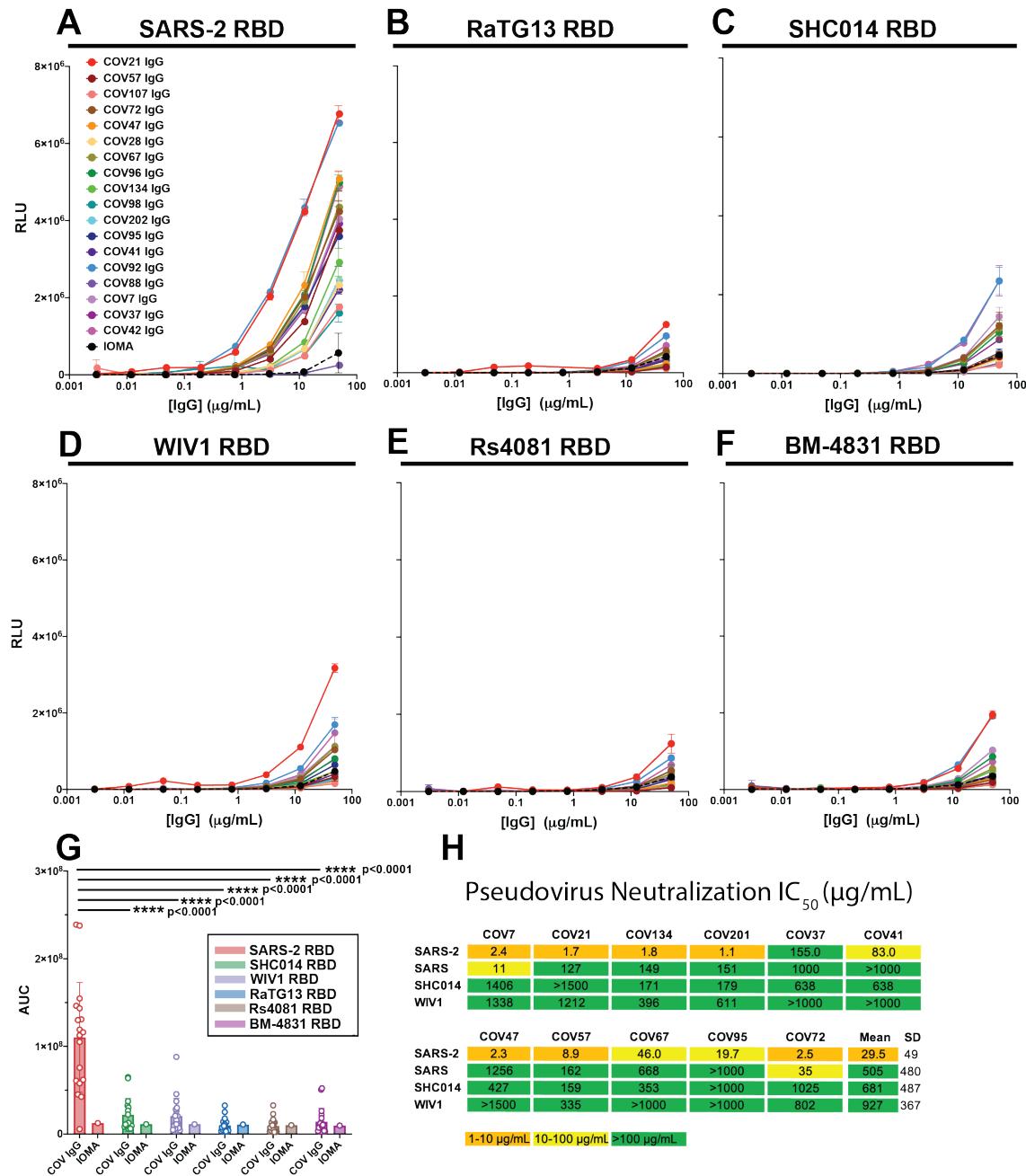


Fig-

ure 4. IgGs from convalescent COVID-19 plasma (18, 24) show little to no cross-reactive responses. (A-F) Plasma IgG responses were evaluated by ELISA (data shown as binding curves with plasma names (18) listed) against RBDs from (A) SARS-2, (B) RaTG13, (C) SHC014, (D) WIV1, (E) Rs4081, and (F) BM-4831. Data points are plotted as the mean and standard deviation of duplicate measurements. IOMA, an anti-HIV-1 IgG (50), was used as a control. (G) ELISA results from panels A-F presented as area under the curve (AUC), where each dot represents one plasma sample, with means and standard deviations represented by rectangles (mean) and horizontal lines (SD). Significant differences between groups linked by horizontal lines are indicated by asterisks and p-values. (H) IC<sub>50</sub> values for pseudotyped neutralization assays using IgGs from COV7, COV21, and COV72 plasmas (18) (evaluated at top concentrations of 1500 µg/mL) against the indicated strains. Mean=arithmetic mean IC<sub>50</sub>; SD=standard deviation.

## Supplementary Materials for

Mosaic nanoparticles elicit cross-reactive immune responses to zoonotic coronaviruses in mice

## Materials and Methods

**Phylogenetic tree.** A sequence alignment of coronavirus RBD domains was made using Clustal Omega (51). A phylogenetic tree was calculated from this amino acid alignment using PhyML 3.0 (52), and a figure of this tree was made using PRESTO (<http://www.atgc-montpellier.fr/presto>).

**Expression of RBD and S proteins.** Mammalian expression vectors encoding the RBDs of SARS-CoV-2 (GenBank MN985325.1; S protein residues 319-539) and SARS-CoV S (GenBank AAP13441.1; residues 318-510) with an N-terminal human IL-2 or Mu phosphatase signal peptide were previously described (47). Expression vectors were constructed similarly for RBDs from the following other sarbecovirus strains: RaTG13-CoV (GenBank QHR63300; S protein residues 319-541), SHC014-CoV (GenBank KC881005; residues 307-524), Rs4081-CoV (GenBank KY417143; S protein residues 310-515), pangolin17-CoV (GenBank QIA48632; residues 317-539), RmYN02-CoV (GSAID EPI\_ISL\_412977; residues 298-503), Rf1-CoV (GenBank DQ412042; residues 310-515), W1V1-CoV (GenBank KF367457; residues 307-528), Yun11-CoV (GenBank JX993988; residues 310-515), BM-4831-CoV (GenBank NC014470; residues 310-530), BtkY72-CoV (GenBank KY352407; residues 309-530). Two versions of each RBD expression vector were made: one including a C-terminal hexahistidine tag (G-HHHHHH) and SpyTag003 (RGVPHIVMVDAYKRYK) (43) (for coupling to SpyCatcher003-mi3) and one with only a hexahistidine tag (for ELISAs). Biotinylated SARS-CoV-2 and Rs4081 RBDs were produced by co-transfection of Avi/His-tagged RBD expression plasmids with an expression plasmid encoding an ER-directed BirA enzyme (kind gift of Michael Anaya, Caltech). RBD proteins were purified from transiently-transfected Expi293F cell

(Gibco) supernatants by nickel affinity and size-exclusion chromatography (47). Peak fractions corresponding to RBDs were identified by SDS-PAGE and then pooled and stored at 4°C. A trimeric SARS-CoV-2 ectodomain with 6P stabilizing mutations (53) was expressed and purified as described (24). Correct folding of the soluble SARS-CoV-2 S trimer was verified by a 3.3 Å cryo-EM structure of a neutralizing antibody complexed with the trimer preparation used for immunizations (24). To prepare fluorochrome-conjugated streptavidin-tetramerized RBDs, biotinylated SARS-2 and Rs4081 RBDs were incubated with streptavidin-APC (eBioscience™) and streptavidin-PE (ThermoFisher), respectively, overnight at 4°C at a 1:1 molar ratio of RBD to streptavidin subunit.

**Preparation of human plasma IgGs.** Plasma samples collected from COVID-19 convalescent and healthy donors are described in (18). Human IgGs were isolated from heat-inactivated plasma samples using 5-mL HiTrap MabSelect SuRe columns (GE Healthcare Life Sciences) as described (24).

**Preparation of RBD-mi3 nanoparticles.** SpyCatcher003-mi3 particles were prepared by purification from BL21 (DE3)-RIPL *E coli* (Agilent) transformed with a pET28a SpyCatcher003-mi3 gene (including an N-terminal 6x-His tag) as described (54). Briefly, cell pellets from transformed bacterial were lysed with a cell disruptor in the presence of 2.0 mM PMSF (Sigma). Lysates were spun at 21,000xg for 30 min, filtered with a 0.2 µm filter, and mi3 particles were isolated by Ni-NTA chromatography using a pre-packed HisTrap™ HP column (GE Healthcare). Eluted particles were concentrated using an Amicon Ultra 15 mL 30K concentrator (MilliporeSigma) and purified by SEC using a HiLoad® 16/600 Superdex® 200 (GE Healthcare) column equilibrated with 25 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.02% NaN<sub>3</sub> (TBS). SpyCatcher-mi3 particles were stored at 4°C and used

for conjugations for up to 1 month after filtering with a 0.2  $\mu$ m filter or spinning at 21,000xg for 10 min.

Purified SpyCatcher003-mi3 was incubated with a 3-fold molar excess (RBD to mi3 subunit) of purified SpyTagged RBD (either a single RBD for making homotypic SARS-CoV-2 RBD particles or an equimolar mixture of four or eight RBDs for making mosaic particles) overnight at room temperature in TBS. Conjugated mi3 particle were separated from free RBDs by SEC on a Superose 6 10/300 column (GE Healthcare) equilibrated with PBS (20 mM sodium phosphate pH 7.5, 150 mM NaCl). Fractions corresponding to conjugated mi3 particles were collected and analyzed by SDS-PAGE. Concentrations of conjugated mi3 particles were determined using a Bio-Rad Protein Assay.

**Immunizations.** Animal procedures and experiments were performed according to protocols approved by the IACUC. Experiments were done using 4-6 week old female Balb/c mice (Charles River Laboratories), with 5 animals each for cohorts immunized with soluble SARS-CoV-2 S or SpyCatcher003-mi3, and 10 animals each for remaining cohorts (Fig 3A). Immunizations were carried out with intraperitoneal (ip) injections of either 5  $\mu$ g of conjugated RBD (calculated as the mass of the RBD, assuming 100% efficiency of conjugation to SpyCatcher003-mi3), 5  $\mu$ g of soluble SARS-CoV-2 S, or 6  $\mu$ g of unconjugated SpyCatcher003-mi3, in 100  $\mu$ L of 50% v/v AddaVax<sup>TM</sup> adjuvant (Invivogen). Animals were boosted 4 weeks after the prime with the same quantity of antigen in adjuvant. Animals were bled every 2 weeks via tail veins, and then euthanized 8 weeks after the prime (Day 56, 57) and bled through cardiac puncture. Blood samples were allowed to clot

at room temperature in MiniCollect® Serum and Plasma Tubes (Greiner), and serum was harvested, preserved in liquid nitrogen, and stored at -80°C until use.

Sera for ELISAs were collected at Day 14 (Prime) and Day 42 (Boost). Sera for neutralization assays were collected at Day 28 (Prime) and Day 56 (Boost) (Fig. 3, fig. S3).

**ELISAs.** 10 µg/ml of a purified RBD (not SpyTagged) in 0.1 M NaHCO<sub>3</sub> pH 9.8 was coated onto Nunc® MaxiSorp™ 384-well plates (Sigma) and stored overnight at 4°C. Plates were washed with Tris-buffered saline with 0.1% Tween 20 (TBS-T) after blocking with 3% bovine serum albumin (BSA) in TBS-T for 1 hr at room temperature. Mouse serum was diluted 1:100 and then serially diluted by 4-fold with TBS-T/3% BSA and added to plates for 3 hr at room temperature. A 1:50,000 dilution of secondary HRP-conjugated goat anti-mouse IgG (Abcam) was added after washing for 1 hr at room temperature. Plates were developed using SuperSignal™ ELISA Femto Maximum Sensitivity Substrate (ThermoFisher) and read at 425 nm. Curves were plotted and integrated to obtain the area under the curve (AUC) using Graphpad Prism 8.3 assuming a one-site binding model with a Hill coefficient (Fig. 3; fig. S3). We also calculated EC<sub>50</sub>s and endpoint titers, which were determined using the dilution that was at or below the mean + 2 x the standard deviation of the plate control (no primary serum added) for ELISA binding data (fig. S3E,F). AUC calculations were used as they better capture changes in maximum binding (55). Statistical significance of titer differences between groups were calculated using Tukey's multiple comparison test using Graphpad Prism 8.3.

**Neutralization assays.** SARS-CoV-2, SARS, WIV1, and SHC014 pseudoviruses based on HIV lentiviral particles were prepared as described (18, 56) using genes encoding S protein sequences lacking C-terminal residues in the cytoplasmic tail: 21 amino acid deletions for SARS-CoV-2, WIV1, and SHC014 and a 19 amino acid deletion for SARS-CoV. IC<sub>50</sub> values derived from this pseudotyped neutralization assay method were shown to quantitatively correlate with results from neutralization assays using authentic SARS-CoV-2 virus (44). For pseudovirus neutralization assays, four-fold serially diluted sera from immunized mice were incubated with a pseudotyped virus for 1 hour at 37°C. After incubation with 293T<sub>ACE2</sub> target cells for 48 hours at 37°C, cells were washed twice with phosphate-buffered saline (PBS) and lysed with Luciferase Cell Culture Lysis 5x reagent (Promega). NanoLuc Luciferase activity in lysates was measured using the Nano-Glo Luciferase Assay System (Promega). Relative luminescence units (RLUs) were normalized to values derived from cells infected with pseudotyped virus in the absence of serum. Half-maximal inhibitory dilutions (ID<sub>50</sub> values) were determined using 4-parameter nonlinear regression in AntibodyDatabase (57). Statistical significance of titer differences between groups was calculated using Tukey's multiple comparison test of ID<sub>50</sub>s converted to log<sup>10</sup> scale using Graphpad Prism 8.3.

**Statistical Analysis.** Comparisons between groups for ELISAs and neutralization assays were calculated with one-way analysis of variance (ANOVA) using Tukey's post hoc test in Prism 9.0 (Graphpad). For correlation analysis between ELISA and neutralization titers, significance (p), Spearman coefficients (r<sub>s</sub>), and linear plots were calculated using Prism 9.0 (Graphpad). Differences were considered significant when p values were less than 0.05. Exact p values are in relevant figure near each corresponding line, with asterisks denoting level of significance (\* denotes 0.01 < p < 0.05, \*\* denotes 0.001 < p < 0.01, \*\*\* denotes 0.0001 < p < 0.001, and \*\*\*\* denotes p < 0.0001).

**Flow cytometry.** B-cell analysis using flow cytometry was carried out as described (54). Briefly, single-cell suspensions were prepared from mouse spleens using mechanical dissociation, and red blood cells were removed using ACK lysing buffer (Gibco). The white blood cell preparation was enriched for IgG+ B-cells using the negative selection protocol in a mouse memory B-cell isolation kit (Miltenyi). The following commercial reagents were used to stain enriched splenocytes: CD4-APC-eFluor 780 (clone: RM4-5), F4/80-APC-eFluor 780 (clone: BM8), CD8a-APC-eFluor 780 (clone: 53-6.7), Ly-6G-APC-eFluor 780 (clone: RB6-8C5), IgM- APC-eFluor 780 (clone: II/41) (Thermo Fisher Scientific), CD19-FITC (clone: 6D5) (Biolegend), IgG1 BV421 (clone: X40) and IgG2 BV421 (clone: R19-15) (BD Bioscience). SARS-2 RBD-APC and Rs4081 RBD-PE for used to identify antigen-specific B-cells. Cell viability was analyzed with Fixable Viability Stain 700 (BD Bioscience). Stained cells were analyzed with a SY3200 Cell Sorter (Sony) configured to detect 6 fluorochromes. 2,000,000 events were collected per sample and analyzed via FlowJo software (TreeStar).

Fig. S1.

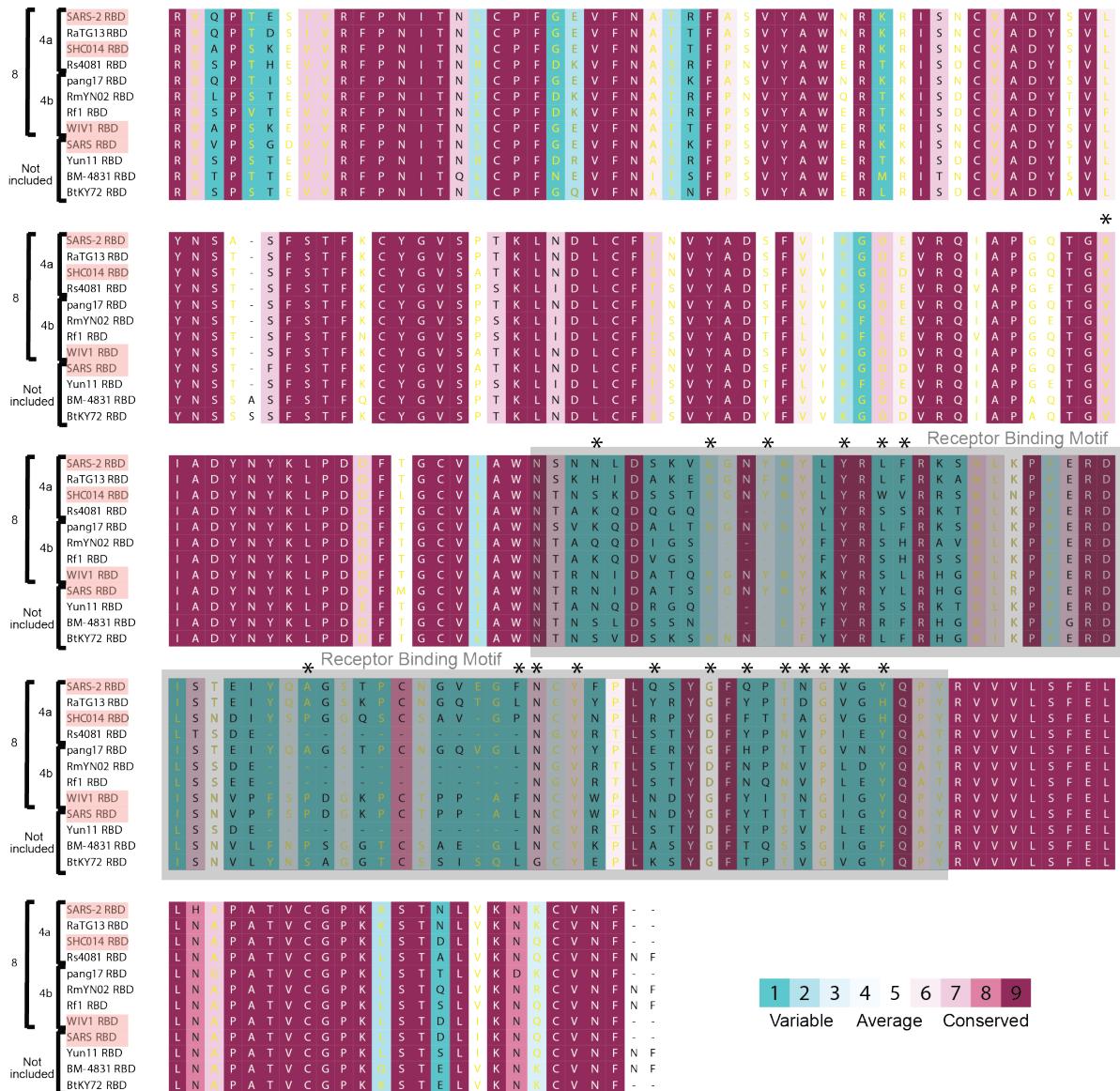
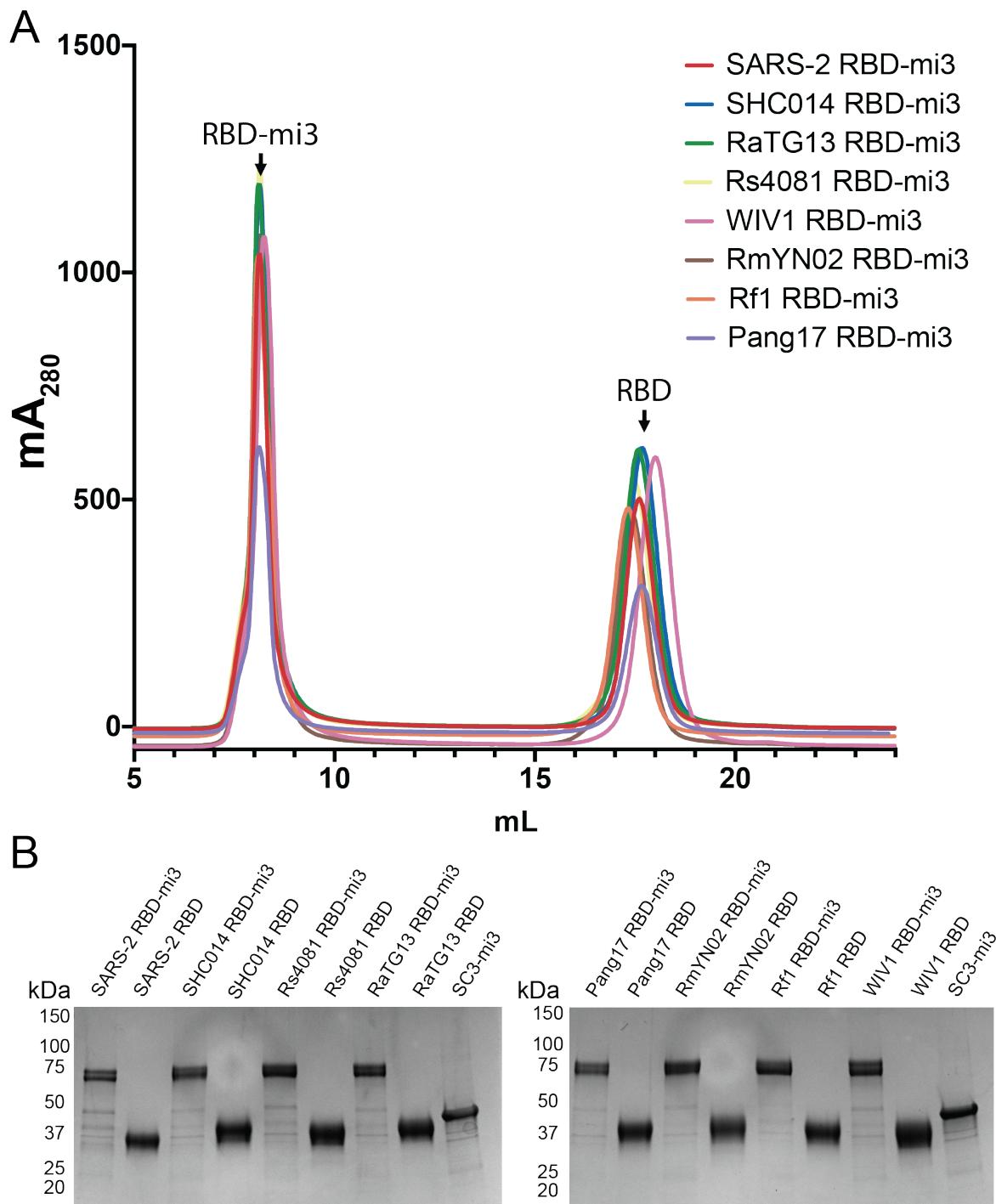
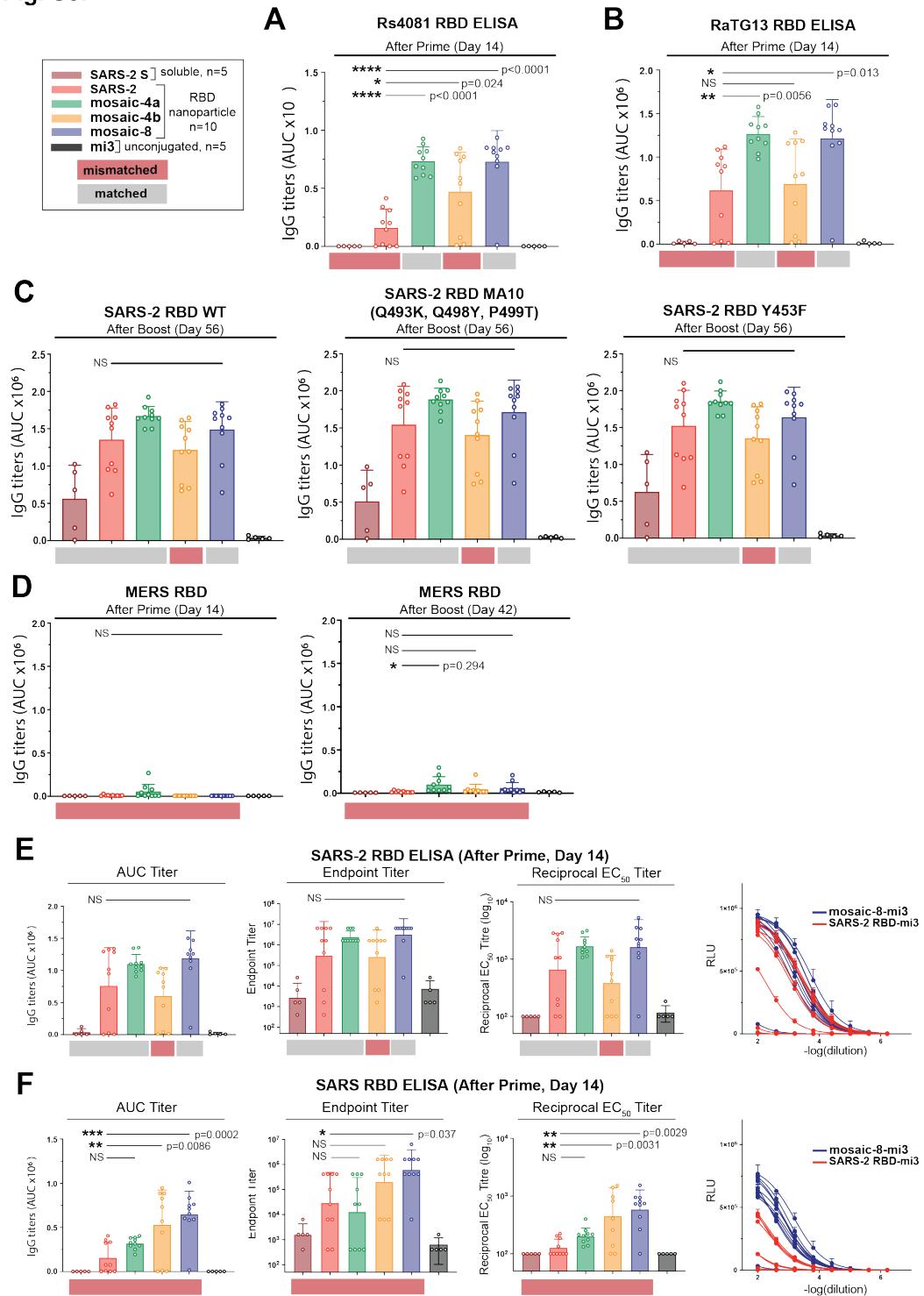


Fig. S1. Alignment of RBD sequences used for making mosaic particles. Sequences shown are for the RBDs of SARS-CoV-2 (SARS-2, GenBank: MN985325.1), RaTG13 (QHR63300), SHC014 (RsSHC014, KC881005), Rs4081 (KY417143), PCoV\_GX-P5L (pang17) (QIA48632), RmYN02 (GSAID EPI\_ISL\_412977), Rf1 (DQ412042), WIV1 (KF367457), SARS-CoV (AAP13441.1), Yun11 (Cp/Yunnan2011, JX993988), BM-4831 (BM48-31/BGR/2008, NC014470), and BtKY72 (KY352407). SARS-2 RBD residues that interact directly with ACE2 (58) are indicated by an asterisk. We note that antibody neutralization by direct binding of ACE2-binding residues does not represent the only mechanism of neutralization for ACE2-tropic viruses. This has been shown for monoclonal human antibodies derived from COVID-19 patients: some neutralizing antibodies do not directly interact with the ACE2-binding site on RBD (for example, class 3 anti-SARS-CoV-2 neutralizing antibodies (24)). Red shading indicates strains known to use ACE2 as a receptor.

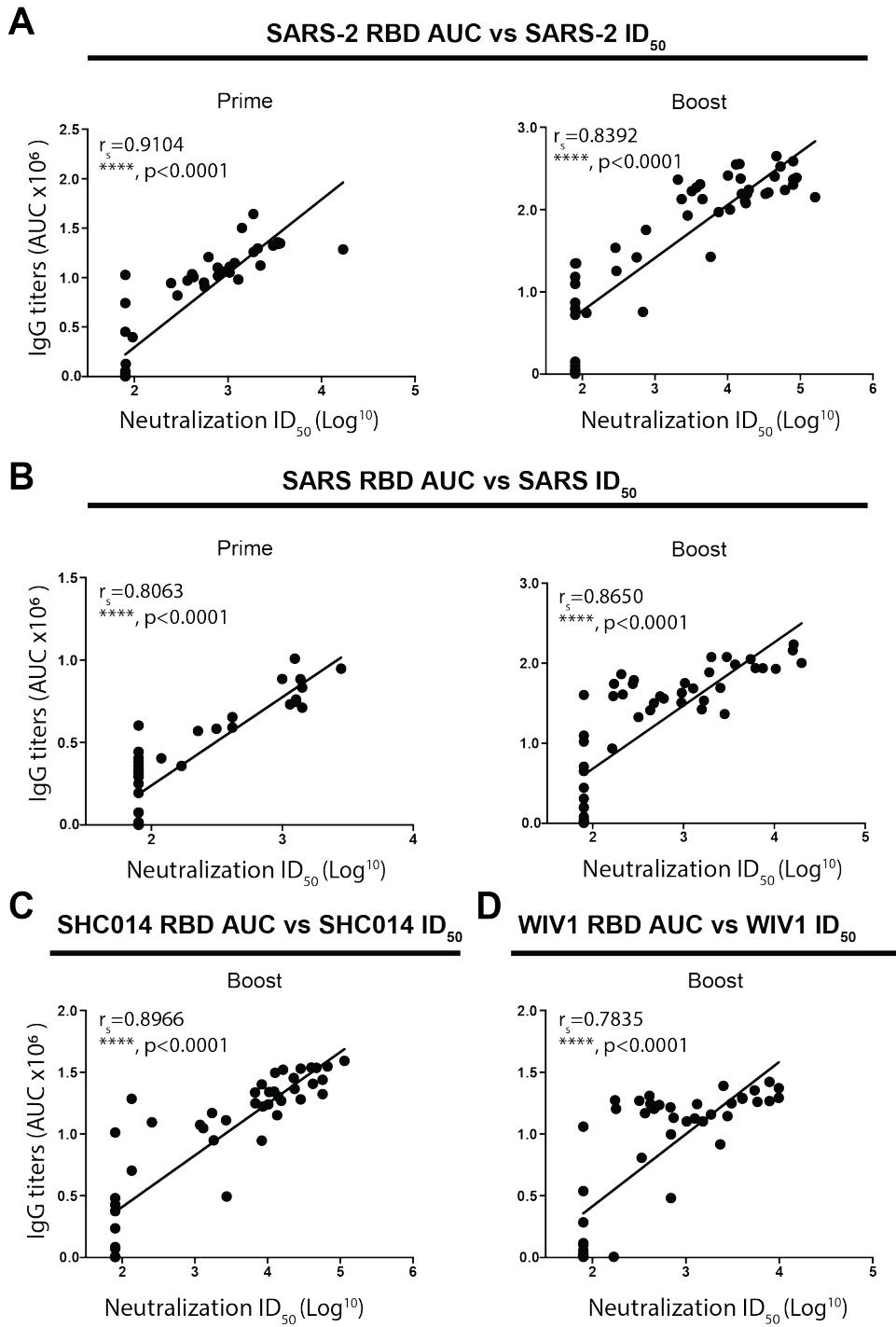
**Fig. S2.**

**Fig. S2.** RBDs from the eight sarbecovirus S proteins conjugate equivalently to SpyCatcher003-mi3, suggesting a statistical mixture of RBDs on mosaic particles. (A) SEC profiles showing separation of RBD nanoparticles and free RBD proteins. (B) Coomassie-stained SDS-PAGE of RBD-coupled nanoparticles, free RBD proteins, and uncoupled SpyCatcher003-mi3 particles (SC3-mi3).

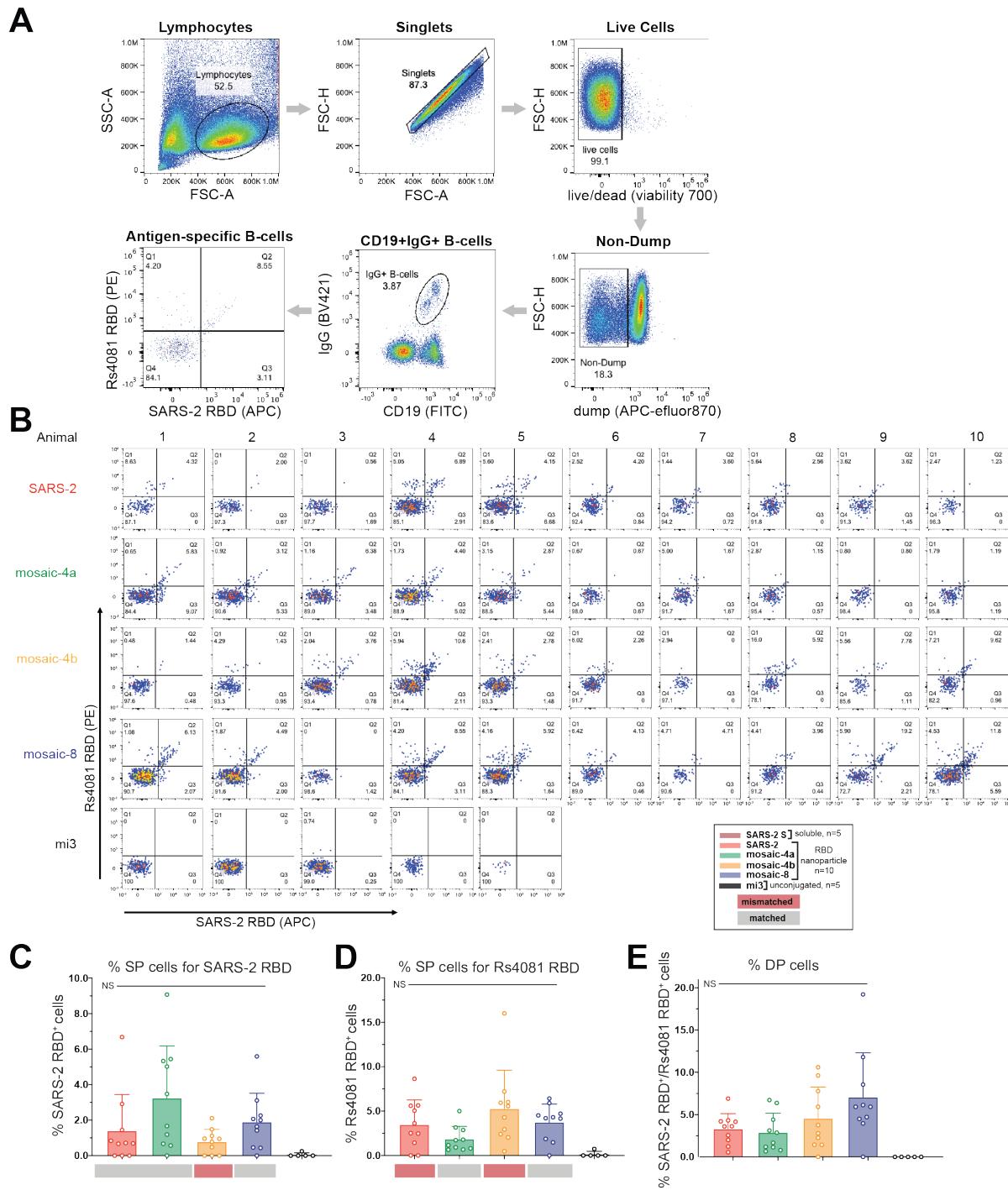
Fig. S3.



**Fig. S3.** Day 14 serum IgG responses to RBDs evaluated by ELISA shown as area under the curve (AUC) from mice immunized with soluble SARS-CoV-2 S trimers (SARS-2 S) or RBDs on nanoparticles (homotypic SARS-2, mosaic-4a, mosaic-4b, mosaic-8, or unconjugated SpyCatcher003-mi3 (mi3)). Each dot represents serum from one animal, with means and standard deviations represented by rectangles (mean) and horizontal lines (SD). RBDs from strains that were not present on an immunized particle or were present on an immunized particle are indicated by red and gray rectangles, respectively, below the ELISA data. Significant differences between groups linked by horizontal lines are indicated by asterisks and p-values. NS=not significant. (A,B) Binding of serum IgGs to (A) Rs4081 and (B) RaTG13 RBDs. (C) Binding of serum IgGs to SARS-2 RBD (left), a triple RBD mutant in a mouse-adapted SARS-CoV-2 (46) that includes substitutions adjacent to the N501Y RBD mutation in an emergent UK SARS-CoV-2 lineage (<https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563>) (middle), and Y453F, the “Danish mink variant” (45) (right). (D) Binding of serum IgGs to RBD from MERS-CoV (a non-ACE2-binding merbecovirus, representing a different subgenus from sarbecoviruses). (E,F) Comparison of ELISA data for serum binding to selected RBDs presented as AUC, endpoint titers, midpoint titers, or binding curves. Day 14 serum IgG responses to (E) SARS-2 or (F) SARS RBDs evaluated by ELISA shown as AUC (left), endpoint titers (middle left), midpoint ( $EC_{50}$ ) titers (middle right), or binding curves (right). For AUC, each dot represents serum from one animal, with means and standard deviations represented by rectangles (mean) and horizontal lines (SD). For endpoint and midpoint titers, each dot represents serum from one animal, with geometric means and geometric standard deviations represented by rectangles (mean) and horizontal lines (SD). Binding curves are shown with data points representing the mean and SD of duplicate measurements fit to a binding model (see Methods) for animals immunized with mosaic-8 and homotypic SARS-2.

**Fig. S4.**

**Fig. S4.** Correlation of ELISA and neutralization titers. Spearman correlation coefficients ( $r_s$ ) and p-values shown for graphs of anti-RBD ELISA titers (AUC) versus pseudovirus neutralization  $ID_{50}$  values; significance indicated as asterisks. (A) SARS-2. (B) SARS. (C) SHC014. (D) WIV1.

**Fig. S5.****Fig S5.** Antigen-specific IgG<sup>+</sup> B-cell analysis of splenocytes isolated from animals immunized with mosaic-RBD nanoparticles. (A) Flow cytometry gating strategy for characterizing RBD-specific IgG<sup>+</sup>

B-cells isolated from splenocytes. Anti-CD4, anti-CD8, anti-F4/80, anti-Ly6G, and anti-IgM were used in the dump to remove T-cells, macrophages, and IgM<sup>+</sup> B-cells. Antigen-specific IgG<sup>+</sup> B-cells were isolated using labeled anti-CD19 and anti-IgG antibodies, and probed for binding RBD with a pair of fluorophore-conjugated RBD trimers (SARS-2 RBD and Rs4081 RBD). (B) Complete flow cytometry analysis for antigen-specific IgG<sup>+</sup> splenocytes isolated from animals immunized with monospecific RBD particles. The 4-way gate shown for each animal separates each population of RBD single-positive and double-positive cells and was used for the % antigen-specific populations shown in panels C-E. Q1 represents the Rs4081 RBD<sup>+</sup> population, Q2 represents the Rs4081 RBD<sup>+</sup> / SARS-2 RBD<sup>+</sup> population, Q3 represents the SARS-2 RBD<sup>+</sup> population, and Q4 represents the RBD<sup>-</sup> population. (C-E) Percent single-positive (SP) and double-positive (DP) cells for the indicated groups. Significant differences between groups linked by horizontal lines are indicated by asterisks and p-values. NS = not significant. (C) Percent SARS-2 RBD<sup>+</sup> B-cells within the IgG<sup>+</sup> B-cell population. (D) Percent Rs4081 RBD<sup>+</sup> B-cells within the IgG<sup>+</sup> B-cell population. (E) Percent SARS-2 RBD<sup>+</sup> / Rs4081 RBD<sup>+</sup> B-cells within the IgG<sup>+</sup> B-cell population.

## References

1. P. Zhou *et al.*, A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270-273 (2020).
2. F. Wu *et al.*, A new coronavirus associated with human respiratory disease in China. *Nature* **579**, 265-269 (2020).
3. E. de Wit, N. van Doremalen, D. Falzarano, V. J. Munster, SARS and MERS: recent insights into emerging coronaviruses. *Nat Rev Microbiol* **14**, 523-534 (2016).
4. W. Li *et al.*, Bats are natural reservoirs of SARS-like coronaviruses. *Science* **310**, 676-679 (2005).
5. N. Wang *et al.*, Serological Evidence of Bat SARS-Related Coronavirus Infection in Humans, China. *Virologica Sinica* **33**, 104-107 (2018).
6. V. D. Menachery *et al.*, A SARS-like cluster of circulating bat coronaviruses shows potential for human emergence. *Nature Medicine* **21**, 1508-1513 (2015).
7. V. D. Menachery *et al.*, SARS-like WIV1-CoV poised for human emergence. *Proceedings of the National Academy of Sciences* **113**, 3048-3053 (2016).
8. W. Li *et al.*, Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* **426**, 450-454 (2003).
9. M. Hoffmann *et al.*, SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* **181**, 271-280 e278 (2020).
10. V. S. Raj *et al.*, Dipeptidyl peptidase 4 is a functional receptor for the emerging human coronavirus-EMC. *Nature* **495**, 251-254 (2013).
11. R. Vlasak, W. Luytjes, W. Spaan, P. Palese, Human and bovine coronaviruses recognize sialic acid-containing receptors similar to those of influenza C viruses. *Proceedings of the National Academy of Sciences* **85**, 4526-4529 (1988).
12. X. Huang *et al.*, Human Coronavirus HKU1 Spike Protein Uses O-Acetylated Sialic Acid as an Attachment Receptor Determinant and Employs Hemagglutinin-Esterase Protein as a Receptor-Destroying Enzyme. *Journal of Virology* **89**, 7202-7213 (2015).
13. T. S. Fung, D. X. Liu, Human Coronavirus: Host-Pathogen Interaction. *Annu Rev Microbiol* **73**, 529-557 (2019).
14. P. J. M. Brouwer *et al.*, Potent neutralizing antibodies from COVID-19 patients define multiple targets of vulnerability. *Science* **369**, 643-650 (2020).
15. Y. Cao *et al.*, Potent neutralizing antibodies against SARS-CoV-2 identified by high-throughput single-cell sequencing of convalescent patients' B cells. *Cell*, (2020).
16. C. Kreer *et al.*, Longitudinal Isolation of Potent Near-Germline SARS-CoV-2-Neutralizing Antibodies from COVID-19 Patients. *Cell*, (2020).
17. L. Liu *et al.*, Potent neutralizing antibodies against multiple epitopes on SARS-CoV-2 spike. *Nature*, (2020).
18. D. F. Robbiani *et al.*, Convergent antibody responses to SARS-CoV-2 in convalescent individuals. *Nature* **584**, 437-442 (2020).
19. R. Shi *et al.*, A human neutralizing antibody targets the receptor-binding site of SARS-CoV-2. *Nature* **584**, 120-124 (2020).
20. S. J. Zost *et al.*, Rapid isolation and profiling of a diverse panel of human monoclonal antibodies targeting the SARS-CoV-2 spike protein. *Nat Med*, (2020).
21. T. F. Rogers *et al.*, Rapid isolation of potent SARS-CoV-2 neutralizing antibodies and protection in a small animal model. *Science*, (2020).

22. E. Seydoux *et al.*, Analysis of a SARS-CoV-2-Infected Individual Reveals Development of Potent Neutralizing Antibodies with Limited Somatic Mutation. *Immunity* **53**, 98-105 e105 (2020).

23. S. J. Zost *et al.*, Potently neutralizing and protective human antibodies against SARS-CoV-2. *Nature* **584**, 443-449 (2020).

24. C. O. Barnes *et al.*, SARS-CoV-2 neutralizing antibody structures inform therapeutic strategies. *Nature*, (2020).

25. D. Pinto *et al.*, Cross-neutralization of SARS-CoV-2 by a human monoclonal SARS-CoV antibody. *Nature* **583**, 290-295 (2020).

26. 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*, (2020).

27. J. López-Sagasta, E. Malito, R. Rappuoli, M. J. Bottomley, Self-assembling protein nanoparticles in the design of vaccines. *Computational and Structural Biotechnology Journal* **14**, 58-68 (2016).

28. M. K. Slifka, I. J. Amanna, Role of Multivalency and Antigenic Threshold in Generating Protective Antibody Responses. *Frontiers in Immunology* **10**, (2019).

29. K. D. Brune, M. Howarth, New Routes and Opportunities for Modular Construction of Particulate Vaccines: Stick, Click, and Glue. *Front Immunol* **9**, 1432 (2018).

30. K. D. Brune *et al.*, Plug-and-Display: decoration of Virus-Like Particles via isopeptide bonds for modular immunization. *Scientific reports* **6**, 19234 (2016).

31. T. U. J. Bruun, A. C. Andersson, S. J. Draper, M. Howarth, Engineering a Rugged Nanoscaffold To Enhance Plug-and-Display Vaccination. *ACS Nano* **12**, 8855-8866 (2018).

32. B. Zakeri *et al.*, Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin. *Proc Natl Acad Sci U S A* **109**, E690-697 (2012).

33. T. K. Tan *et al.*, A COVID-19 vaccine candidate 1 using SpyCatcher multimerization of the SARS-CoV-2 spike protein receptor-binding domain induces potent neutralising antibody responses. *bioRxiv*, (2020).

34. B. Zhang *et al.*, A platform incorporating trimeric antigens into self-assembling nanoparticles reveals SARS-CoV-2-spike nanoparticles to elicit substantially higher neutralizing responses than spike alone. *Scientific reports* **10**, (2020).

35. C. Drosten *et al.*, Discovery of a rich gene pool of bat SARS-related coronaviruses provides new insights into the origin of SARS coronavirus. *PLOS Pathogens* **13**, e1006698 (2017).

36. R. Rahikainen *et al.*, Overcoming Symmetry Mismatch in Vaccine Nanoassembly through Spontaneous Amidation. *Angewandte Chemie International Edition*, (2020).

37. M. Kanekiyo *et al.*, Mosaic nanoparticle display of diverse influenza virus hemagglutinins elicits broad B cell responses. *Nat Immunol* **20**, 362-372 (2019).

38. M. Letko, A. Marzi, V. Munster, Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. *Nature Microbiology* **5**, 562-569 (2020).

39. T. T.-Y. Lam *et al.*, Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins. *Nature* **583**, 282-285 (2020).

40. K. G. Andersen, A. Rambaut, W. I. Lipkin, E. C. Holmes, R. F. Garry, The proximal origin of SARS-CoV-2. *Nature Medicine* **26**, 450-452 (2020).

41. H. Zhou *et al.*, A Novel Bat Coronavirus Closely Related to SARS-CoV-2 Contains Natural Insertions at the S1/S2 Cleavage Site of the Spike Protein. *Current Biology* **30**, 2196-2203.e2193 (2020).

42. Y. Tao, S. Tong, K. M. Stedman, Complete Genome Sequence of a Severe Acute Respiratory Syndrome-Related Coronavirus from Kenyan Bats. *Microbiology Resource Announcements* **8**, (2019).

43. A. H. Keeble *et al.*, Approaching infinite affinity through engineering of peptide–protein interaction. *Proceedings of the National Academy of Sciences* **116**, 26523-26533 (2019).

44. F. Schmidt *et al.*, Measuring SARS-CoV-2 neutralizing antibody activity using pseudotyped and chimeric viruses. *Journal of Experimental Medicine* **217**, (2020).

45. L. van Dorp *et al.*, Recurrent mutations in SARS-CoV-2 genomes isolated from mink point to rapid host-adaptation. *bioRxiv*, (2020).

46. S. R. Leist *et al.*, A Mouse-Adapted SARS-CoV-2 Induces Acute Lung Injury and Mortality in Standard Laboratory Mice. *Cell* **183**, 1070-1085.e1012 (2020).

47. C. O. Barnes *et al.*, Structures of Human Antibodies Bound to SARS-CoV-2 Spike Reveal Common Epitopes and Recurrent Features of Antibodies. *Cell* **182**, 828-842 e816 (2020).

48. A. C. Walls *et al.*, Elicitation of Potent Neutralizing Antibody Responses by Designed Protein Nanoparticle Vaccines for SARS-CoV-2. *Cell* **183**, 1367-1382.e1317 (2020).

49. M. Landau *et al.*, ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res* **33**, W299-302 (2005).

50. H. B. Gristick *et al.*, Natively glycosylated HIV-1 Env structure reveals new mode for antibody recognition of the CD4-binding site. *Nat Struct Mol Biol* **23**, 906-915 (2016).

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

52. S. Guindon *et al.*, New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* **59**, 307-321 (2010).

53. C.-L. Hsieh *et al.*, Structure-based design of prefusion-stabilized SARS-CoV-2 spikes. *Science* **369**, 1501-1505 (2020).

54. A. A. Cohen *et al.*, Construction, characterization, and immunization of nanoparticles that display a diverse array of influenza HA trimers. *bioRxiv*, (2020).

55. D. Angeletti *et al.*, Defining B cell immunodominance to viruses. *Nature Immunology* **18**, 456-463 (2017).

56. K. H. D. Crawford *et al.*, Protocol and Reagents for Pseudotyping Lentiviral Particles with SARS-CoV-2 Spike Protein for Neutralization Assays. *Viruses* **12**, 513 (2020).

57. A. P. West, Jr. *et al.*, Computational analysis of anti-HIV-1 antibody neutralization panel data to identify potential functional epitope residues. *Proc Natl Acad Sci U S A* **110**, 10598-10603 (2013).

58.T. N. Starr *et al.*, Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding. *Cell* **182**, 1295-1310.e1220 (2020).

Appendix A: Broad cross-reactivity across sarbecoviruses exhibited by a subset of COVID-19 donor-derived neutralizing antibodies

This work describes the characterization of two antibodies isolated from COVID-19 patients. These antibodies are both cross-reactive and cross-neutralizing against a panel of sarbecoviruses. Structural studies reveal binding to a highly conserved epitope on the RBD with blocking of ACE2 receptor binding being the proposed mechanism of neutralization due to recognition of residues that lie closer to the ACE2 binding motif.

Claudia A. Jette carried out the crystallography studies as well as designed of some of the reagents including the bispecific antibodies. Christopher Barnes this the single particle Cryo-EM studies. My contribution to this work was as the co-lead researcher of the project. I helped conceive and design the cross-reactive ELISA and neutralization experiments to both sarbecovirus RBDs as well as variants of concern, prepared the reagents, and analyzed the data.

This work was submitted to a preprint as:

Jette, C.A., **Cohen, A.A.**, Gnanapragasam, P.N.P., Muecksch, F., Lee, Y.E., Huey-Tubman, K.E., Schmidt, F., Hatzioannou, T., Bienasz, P.D., Nussenzweig, M.C., et al. (2021). Broad cross-reactivity across sarbecoviruses exhibited by a subset of COVID-19 donor-derived neutralizing antibodies. *Biorxiv* 2021.04.23.441195.

**Summary**

Many anti-SARS-CoV-2 neutralizing antibodies target the ACE2-binding site on viral spike receptor-binding domains (RBDs). The most potent antibodies recognize exposed variable epitopes, often rendering them ineffective against other sarbecoviruses and SARS-CoV-2 variants. Class 4 anti-RBD antibodies against a less-exposed, but more-conserved, cryptic epitope could recognize newly-emergent zoonotic sarbecoviruses and variants, but usually show only weak neutralization potencies. We characterized two class 4 anti-RBD antibodies derived from COVID-19 donors that exhibited broad recognition and potent neutralization of zoonotic coronavirus and SARS-CoV-2 variants. C118-RBD and C022-RBD structures revealed CDRH3 mainchain H-bond interactions that extended an RBD  $\beta$ -sheet, thus reducing sensitivity to RBD sidechain changes, and epitopes that extended from the cryptic epitope to occlude ACE2 binding. A C118-spike trimer structure revealed rotated RBDs to allow cryptic epitope access and the potential for intra-spike crosslinking to increase avidity. These studies facilitate vaccine design and illustrate advantages of class 4 RBD-binding antibody therapeutics.

**Key words**

Cryo-EM; Coronavirus; Neutralizing antibody; receptor-binding domain; Sarbecovirus; SARS-CoV-2; Spike trimer; Structural biology; Virology; X-ray crystallography.

## Introduction

The current SARS-CoV-2 pandemic is a crisis of immediate global concern, but two other zoonotic betacoronaviruses, SARS-CoV and MERS-CoV (Middle East Respiratory Syndrome), also resulted in epidemics within the last 20 years (de Wit et al., 2016). All three viruses likely originated in bats (Li et al., 2005; Zhou et al., 2021), with SARS-CoV and MERS-CoV having adapted to intermediary animal hosts, most likely palm civets (Song et al., 2005) and dromedary camels (Haagmans et al., 2014), respectively, prior to infection of humans. Serological surveys of people living near caves where bats carry diverse coronaviruses suggests direct transmission of SARS-CoV-like viruses (Wang et al., 2018), raising the possibility of future outbreaks resulting from human infection with SARS-like betacoronaviruses (sarbecoviruses).

Coronaviruses encode a trimeric spike glycoprotein (S) that serves as the machinery for fusing the viral and host cell membranes (Fung and Liu, 2019). The first step in fusion is contact of S with a host receptor. The receptor-binding domains (RBDs) at the apex of the S trimers of SARS-CoV-2, SARS-CoV, HCoV-NL63, and some animal coronaviruses utilize angiotensin-converting enzyme 2 (ACE2) as their receptor (Hoffmann et al., 2020; Li et al., 2003; Zhou et al., 2020b). RBDs can adopt either 'down' or 'up' conformations, with ACE2 binding only possible to RBDs in an 'up' conformation (Kirchdoerfer et al., 2016; Li et al., 2019; Walls et al., 2020; Walls et al., 2016; Wrapp et al., 2020; Yuan et al., 2017). A phylogenetic tree of the relationship between coronavirus S protein RBDs shows that sarbecovirus RBDs form a separate branch (Figure 1A).

Consistent with their obligate role in viral entry, sarbecovirus S trimers are the primary targets of neutralizing antibodies (Brouwer et al., 2020; Cao et al., 2020; Fung and Liu, 2019; Kreer et al., 2020; Liu et al., 2020b; Robbiani et al., 2020; Rogers et al., 2020; Seydoux et al., 2020; Shi et al., 2020; Zost et al., 2020b), with many focusing on the RBD (Barnes et al., 2020a; Barnes et al., 2020b; Brouwer et al., 2020; Cao et al., 2020; Kreer et al., 2020; Liu et al., 2020b; Pinto et al., 2020; Robbiani et al., 2020; Rogers et al., 2020; Seydoux et al., 2020; Zost et al., 2020a). Structural analysis of the binding epitopes of anti-SARS-CoV-2 RBD antibodies enabled their classification into four initial categories: class 1, derived from *VH3-53/VH3-63* germlines and including a short heavy chain complementarity determining region 3 (CDRH3) that bind an epitope overlapping with the ACE2 binding site and only recognize 'up' RBDs; class 2, whose epitope also overlaps with the ACE2 binding site, but which can bind to both 'up' and 'down' RBD conformations; class 3, which bind to the opposite side of 'up' and 'down' RBDs adjacent to an N-glycan attached to residue N343; and class 4, which are often weakly neutralizing antibodies that target a cryptic epitope facing the interior of the spike protein on 'up' RBDs (Barnes et al., 2020a) (Figure S1; Supplemental Movie 1).

Potent anti-SARS-CoV-2 neutralizing antibodies are typically class 1 or class 2 anti-RBD antibodies that block the ACE2 binding site (Barnes et al., 2020a; Dejnirattisai et al., 2021; Lee et al., 2021; Liu et al., 2020b; Tortorici, 2020). Since class 1 and class 2 RBD epitopes are not well conserved (Figure 1B), antibodies in these classes are unlikely to strongly cross-react across sarbecovirus RBDs. However, an *in vitro*-selected variant of an ACE2 blocking antibody isolated from a SARS-infected survivor exhibited increased cross-reactive properties, showing neutralization of SARS-CoV-2 and other betacoronaviruses (Rappazzo et al., 2021). In general, however, as isolated from infected donors, class 3 and class 4 RBD-binding antibodies are better prospects for neutralizing across multiple strains and thereby potentially protecting against emergent sarbecoviruses. Indeed, S309, a class 3 anti-RBD antibody isolated from a SARS-CoV-infected donor, demonstrated cross-

reactive neutralization of SARS-CoV-2 (Pinto et al., 2020). Furthermore, reports of class 4 human antibodies that exhibit cross-reactive binding and neutralization amongst sarbecoviruses (Liu et al., 2020a; Starr et al., 2021a) suggest that further investigation of antibodies from COVID-19 convalescent donors could lead to discoveries of potent and broadly cross-reactive class 4 antibodies that recognize the highly-conserved, 'cryptic' RBD epitope.

Here we investigated C118 and C022, two class 4 human antibodies isolated from COVID-19 donors (Robbiani et al., 2020) that show breadth of binding and neutralization across sarbecoviruses and SARS-CoV-2 variants of concern. We report crystal structures of C118 complexed with SARS RBD and C022 complexed with SARS-CoV-2 RBD, which revealed interactions with a conserved portion of the RBD in common with interactions of previously-described cross-reactive but more weakly-neutralizing class 4 antibodies; e.g., CR3022 (Huo et al., 2020; Yuan et al., 2020a; Yuan et al., 2020b), and EY6A (Zhou et al., 2020a). Unlike these class 4 anti-RBD antibodies, C118 and C022 also occlude portions of the ACE2 binding site to facilitate more potent neutralization. A single-particle cryo-EM structure of a C118-S trimer complex demonstrated binding of C118 to an intact trimer, revealing an S configuration with increased separation between the RBDs than found in class 1-3 Fab-S or ACE2-S trimer structures, and revealed the potential for intra-spike crosslinking. These results define a cross-reactive class 4-like epitope on sarbecovirus RBDs that can be targeted in vaccine design and illustrate a mechanism by which the cryptic RBD epitope can be accessed on intact CoV S trimers.

## Results

### C022 and C118 IgGs recognize and neutralize diverse sarbecoviruses, including SARS-CoV-2 variants

From a survey to identify cross-reactive monoclonal antibodies isolated from SARS-CoV-2-infected donors from the New York area (Robbiani et al., 2020), we found antibodies isolated from different donors, C118 (VH3-30/VL4-69-encoded) and C022 (VH4-39/VK1-5-encoded), that recognized a diverse panel of 12 sarbecovirus RBDs spanning clades 1, 1/2, 2 and 3 (Figure 1). As evaluated by enzyme-linked immunosorbent assay (ELISA), C118 bound to RBDs from all sarbecoviruses tested, and C022 bound to all but two RBDs, similar to the class 4 anti-RBD antibody CR3022 (Figure 1C). By comparison, the cross-reactive class 3 anti-SARS RBD antibody S309 (Pinto et al., 2020) recognized half of the set of sarbecovirus RBDs, and C144, a more potent SARS-CoV-2 class 2 neutralizing antibody (Robbiani et al., 2020), bound to the SARS-CoV-2 RBD but not to RBDs from the other 11 sarbecovirus strains (Figure 1C).

To further define the C022 and C118 antibody epitopes, we evaluated binding of C118 and C022 to a panel of RBDs with mutations chosen from circulating variants that conferred resistance to one or more classes of anti-RBD antibodies (Li et al., 2020; Starr et al., 2021b; Weisblum et al., 2020). We also assessed binding to RBD substitutions identified in the B.1.1.7 and B.1.351 SARS-CoV-2 variants of concern (Rambaut et al., 2020; Tegally et al., 2020), and to mutations in the MA10 mouse-adapted SARS-CoV-2 virus (Leist et al., 2020). Relative to wild-type RBD, C118, C022, CR3022 and S309 demonstrated a similar binding profile with respect to the RBD substitutions tested and exhibited a broader range of binding to the RBD mutants than did the more potent class

2 C144 antibody (Figure 1C and Figure S2A). Collectively, the ELISA binding data suggested that C022 and C118 recognize a highly-conserved epitope and are therefore likely to be class 4 anti-RBD antibodies.

We next measured neutralization potencies using an in vitro pseudovirus-based assay that quantitatively correlates with authentic virus neutralization (Schmidt et al., 2020) to evaluate SARS-CoV-2, SARS-CoV-2 RBD mutants, SARS-CoV-2 variants (Annavajhala et al., 2021; Faria et al., 2021; Rambaut et al., 2020; Tegally et al., 2020; Voloch et al., 2020; West et al., 2021; Zhang et al., 2021), and sarbecovirus strains known to infect human ACE2-expressing target cells (SARS-CoV-2, SARS-CoV, WIV1, SHC104, WIV16, Pangolin GD and Pangolin GX) (Figure 1D,E and Figure S2B-D). Against a panel of SARS-CoV-2 pseudotyped viruses harboring single amino acid RBD substitutions, C118 and C022 neutralized all viruses with potencies similar to 'wt' SARS-CoV-2, consistent with the results obtained in ELISA binding assays (S gene with D614 residue; GenBank: NC\_045512) (Figure S2). For comparisons with SARS-CoV-2 variants of concern, the S gene we used to make 'wt' SARS-CoV-2 pseudovirus included the D614G substitution in the context of the Wuhan-1 spike (Korber et al., 2020), resulting in a 2-4-fold reduction in IC<sub>50</sub>s for C022 and C118 antibodies (Figure 1E).

We found that C118 and C022 IgGs neutralized all four SARS-CoV-2 variants and all ACE2-tropic sarbecoviruses with 50% inhibitory concentrations (IC<sub>50</sub> values) of <1 µg/mL, with the exception of C118, which inhibited SARS-CoV-pseudotyped viruses less efficiently (IC<sub>50</sub> = ~4.5 µg/mL) (Figure 1D,E and Figure S2B-D). By contrast, the class 4 anti-RBD antibody CR3022 showed weak or no neutralization against the majority of pseudoviruses tested, with the exception of SARS-CoV (IC<sub>50</sub> ~1.1 µg/mL) and WIV1 (IC<sub>50</sub> ~0.6 µg/mL). The class 3 S309 antibody showed strong neutralization potencies (IC<sub>50</sub>s between 16 ng/mL and 120 ng/mL) against all viruses with the exceptions of the B.1.1.7 SARS-CoV-2 variant of concern and SHC014. The class 2 anti-RBD antibody C144 was highly potent against SARS-CoV-2 and the B.1.1.7 and B.1.429 variants (IC<sub>50</sub>s between 1 ng/mL and 2 ng/mL), but did not neutralize the other SARS-CoV-2 variants or sarbecoviruses. Taken together, of the IgGs evaluated, C118 and C022 exhibited the greatest breadth of sarbecovirus neutralization (Figure 1E and Figure S2), consistent with their broad cross-reactive binding profile demonstrated by ELISA (Figure 1C and Figure S2A).

### **Crystal structures of C022-RBD and C118-RBD reveal class 4 RBD interactions and conservation of epitope residues**

To understand the mechanism underlying the breadth of neutralization of C022 and C118, we solved structures of complexes between C118 Fab bound to SARS-CoV RBD and C022 bound to SARS-CoV-2 RBD to resolutions of 2.7 Å and 3.2 Å, respectively, chosen based on which complexes formed well-ordered crystals (Figure 2A,B and Table S1).

The C118-RBD and C022-RBD structures showed that both Fabs recognize an epitope that is highly-conserved among sarbecoviruses at the base of the RBD (Figure 1B), which is exposed only in 'up' RBD conformations as first described for the class 4 RBD-binding antibodies CR3022 (Huo et al., 2020; Yuan et al., 2020a; Yuan et al., 2020b) and EY6A (Zhou et al., 2020a). C022 and C118 use four of six complementarity-determining region (CDR) loops to interact with an epitope that extends towards the RBD ridge near the ACE2 binding site, and in the case of C022, includes an

overlapping interacting residue (K417<sub>RBD</sub>) (Figure 2C,D). In both structures, CDRH3 loops, CDRL2 loops, and portions of FWRL3 mediate the majority of RBD contacts and establish extensive polar and van der Waals interactions with RBD residues (Figure 2C,D), accounting for 71% of epitope buried surface area (BSA) on the RBD for the C022-RBD and C118-RBD structures, respectively (Table S2). No contacts were made in either complex with the N343<sub>RBD</sub> N-glycan (SARS-CoV-2 S numbering). SARS-CoV contains an additional potential *N*-linked glycosylation site at N357<sub>RBD</sub> (SARS-CoV S numbering), which if glycosylated, would not be contacted by C118, a favorable feature for cross-reactive recognition given that this potential *N*-linked glycosylation site is conserved in all S protein sequences except for SARS-CoV-2 (Figure 2A).

Overlaying the RBDs of our Fab-RBD structures with the RBD of the ACE2-RBD structure (PDB 6M0J) showed that the binding poses of both C118 and C022 placed the V<sub>L</sub> of each Fab in a position that would clash with concurrent ACE2 binding, in contrast to the CR3022 and EY6A binding poses (Figure 2E). This binding orientation would sterically prevent RBD-ACE2 interactions, as has been suggested for other class 4 anti-RBD antibodies (Liu et al., 2020a; Piccoli et al., 2020a). To verify direct competition with ACE2, we conducted a competition experiment using surface plasmon resonance (SPR). SARS-CoV-2 RBD was coupled to a biosensor chip, an RBD-binding IgG was injected, and then soluble ACE2 was injected over the RBD-IgG complex. C118, C022, and C144 IgGs all inhibited binding of ACE2 did not bind to the IgG-RBD complex. In contrast, ACE2 bound the CR3022-RBD complex (Figure S3). These results are consistent with competition for C118, C022, and C144 for ACE2 binding to RBD, but no competition for CR3022, suggesting a primary neutralization mechanism for C022 and C118 that prevents spike attachment to host cell ACE2 receptors.

### Features of C118 and C022 recognition of the class 4 epitope

Class 4 RBD-binding antibodies contact a common epitope at the base of the RBD that is distant from the ACE2-binding site (Figure 3A). The epitopes of three class 4 antibodies, C118, C022, and COVA1-16, also includes a patch reaching towards the ridge on the left side of the RBD as depicted in Figure 3A.

To compare the C118 and C022 epitopes with epitopes of other class 4 anti-RBD antibodies, we analyzed RBD residues contacted by C118, C022, COVA1-16, and CR3022 on aligned sequences of sarbecovirus RBDs (Figure 3B). Sequence conservation among sarbecoviruses at the C022 and C118 epitopes involves a majority of residues that are strictly-conserved or conservatively-substituted between SARS-CoV-2 and other RBDs (Figure 3B), likely explaining the broad cross-reactivity observed for these antibodies (Figure 1C). Comparison of the C118 and C022 epitopes showed a majority of recognized RBD residues are shared between the two antibodies (70% of C118 epitope also contacted by C022) (Figure 3B). CR3022 contacted a similar number of residues as C118 and C022, including the conserved patch at the RBD base (Figure 3B,C); however, a region from 404-417<sub>RBD</sub> that comprises an unstructured loop and the  $\alpha$ 4 helix above an internal RBD  $\beta$ -sheet contained only a single CR3022 contact residue (R408<sub>RBD</sub>) and was not contacted by antibodies EY6A, S2A4 and S304; whereas C118, C022, and COVA1-16 showed contacts with this region (Figure 3B,D).

The  $\alpha 4$  helix is proximal to the ACE2 receptor-binding motif and has less sequence conservation across the 12 sarbecoviruses (Figure 3B). To accommodate binding in this region, C118 uses insertions in its FWRL3 (54B-56<sub>LC</sub>) to form a  $\beta$ -strand adjacent to the  $\alpha 4$  helix, establishing both side chain and backbone interactions (Figure 3E – left panel). C022 showed similar binding in this region but used non-contiguous CDRH1, CDRH3, and CDRL2 loops (Figure 3E – right panel). C022 contacts were located more to the C-terminal end of the  $\alpha 4$  helix than the C118 contacts and encompassed the disordered RBD loop that includes the ACE2-interacting residue K417<sub>RBD</sub> (K404<sub>RBD</sub> in SARS-CoV) (Lan et al., 2020) (Figure 3E – right panel). C022 buried more surface area on RBD in this region than C118 (323 $\text{\AA}^2$  vs 150 $\text{\AA}^2$ ). Four of eight and five of nine RBD contacts for C118 and C022, respectively, were fully conserved among sarbecoviruses (Figure 3B), suggesting that interactions in this region may be possible with other sarbecoviruses. In particular, the conserved residue R408<sub>RBD</sub> (R395<sub>RBD</sub> in SARS-CoV) was contacted by both antibodies and alone was responsible for 94 $\text{\AA}^2$  and 95 $\text{\AA}^2$  of BSA buried on the RBDs for C118 and C022, respectively. Despite both C118 and C022 engaging the  $\alpha 4$  helix and residue R408<sub>RBD</sub>, mutations at this position known to affect class 1 and class 4 anti-RBD antibodies (Greaney et al., 2021) had no effect on these antibodies (Figure S2A). Overall, engagement of the  $\alpha 4$  helix region provided 16% (C118) and 36% (C022) of the BSA buried on RBD, and extended their epitopes past the cryptic epitope to bind adjacent to or overlapping with the ACE2 binding site.

### Shared features of the C022 and COVA1-16 class 4 anti-RBD antibodies

The C022 epitope on RBD closely resembles the epitope of COVA1-16 (Figure 3A,B), a class 4 antibody isolated from a SARS-CoV-2 convalescent donor derived from VH1-46/VK1-33 V-gene segments (Brouwer et al., 2020) (Figure S4). Yet, COVA1-16 weakly neutralized (>1 $\mu\text{g/mL}$ ) SARS-CoV and SARS-CoV-2 variants (Liu et al., 2021), in contrast with C022 neutralization (Figure 1E). After superimposing the RBDs from crystal structures of SARS-CoV-2 RBD complexed with COVA1-16 (PDB 7JMW) and C022 (this study), the V<sub>H</sub>-V<sub>L</sub> domains of the bound Fabs were related by an root mean square deviation (RMSD) of 1.3 $\text{\AA}$  (235 C $\alpha$  atoms), with the majority of conformational differences occurring in the CDRH1 and CDRH2 loops (Figure S5A). Despite being derived from different V gene segments (which would affect their VH gene segment-encoded CDRH1 and CDRH2 loops), C022 and COVA1-16 recognized similar epitopes, contacting a common set of 23 RBD residues, which included COVA1-16 interactions in common with C022 interactions with the RBD  $\alpha 4$  helix (Figure 3B).

While C022 and COVA1-16 share a generally similar mode of binding, there are differences in interactions of residues encoded within their different V<sub>H</sub> gene segments (i.e., their CDRH1 and CDRH2 loops) (Figure S5B). For example, the C022 contact with T430<sub>RBD</sub> was part of an extensive clasp made by an interaction between the C022 CDRH1 residue R33<sub>HC</sub> with backbone carbonyls of D427<sub>RBD</sub>, D428<sub>RBD</sub>, and F429<sub>RBD</sub> and with the sidechain of T430<sub>RBD</sub> (Figure S5C). Two of the same RBD residues (D427<sub>RBD</sub> and F429<sub>RBD</sub>) interacted with an arginine from COVA1-16, but this arginine (R100B<sub>HC</sub>) is located at the base of the CDRH3 loop rather than within CDRH1, as is the case with C022 R33<sub>HC</sub>. The larger separation distance from the RBD of COVA1-16 R100B<sub>HC</sub> allowed it to form a sidechain-backbone H-bond with D427<sub>RBD</sub> similar to a sidechain-backbone H-bond involving C022 R33<sub>HC</sub> and D428<sub>RBD</sub>, but precluded interactions with D428<sub>RBD</sub> and T430<sub>RBD</sub> (Figure S5D). In

addition, the COVA1-16 CDRH1 was shorter than the C022 CDRH1 (7 versus 9 residues) (Figure S4A), and was shifted away from the RBD relative to the C022 CDRH1. These differences, in addition to fewer LC interactions by COVA1-16, resulted in less total BSA for COVA1-16 relative to C022 (1607 Å<sup>2</sup> vs 1875 Å<sup>2</sup>, respectively) despite similar contributions from CDRH3 loops (Table S2).

### Interactions with RBD main chain atoms facilitate recognition of diverse RBDs

The paratopes of both C118 and C022 were dominated by their long CDRH3 loops (20 and 21 residues, respectively) (Figure 4A,B and Figure S4), which make up ~half of the buried surface areas (BSAs) of each paratope (461 Å<sup>2</sup> of 1020 Å<sup>2</sup> for C118 and 537 Å<sup>2</sup> of 969 Å<sup>2</sup> for C022) (Table S2). The C118 and C022 CDRH3s comprise two anti-parallel β-strands that extend a largely internal RBD β-sheet (β-strands β1-β4 and β7) through main chain H-bonds between the RBD β2 strand (377-379<sub>RBD</sub>) and the first CDRH3 β-strand (CDRH3 residues 97-99 (C118) or 100-100B (C022)) (Figure 4A,B). A similar feature is also seen in the structure of the COVA1-16-RBD complex (Liu et al., 2020a), which shares a nearly identical CDRH3 sequence with C022 (Figure S5E).

C118 and C022 form extensive backbone interactions with RBD, with 10 and 9 H-bonds formed with the backbone of RBD, respectively. Extensive backbone interactions in the C118 and C022 epitopes could contribute to their breadth of binding and neutralization across sarbecoviruses, as backbone interactions would facilitate binding despite side chain substitutions, which are rare across the RBD sequences listed (Figure 3B), but could occur in other CoV RBDs. For example, the backbone H-bonds between the CDRH3s of C118 and C022 with the RBD β2 strand allow for binding despite substitution at position K378<sub>RBD</sub> (K365<sub>RBD</sub> in SARS-CoV) (Figure S2A and Figure 4C). By contrast, the class 4 antibody CR3022 uses side chain interactions (potential electrostatic interactions between D54<sub>HC</sub> and E56<sub>HC</sub> and K378<sub>RBD</sub>); thus CR3022 is sensitive to mutation at K378N<sub>RBD</sub> (Figure S2A). This is consistent with CR3022 not binding to Rf1-CoV RBD (Figure 1C), which contains an asparagine at the equivalent position to SARS-CoV-2 K378<sub>RBD</sub> (Figure 3B), whereas C118 and C022 binding to Rf1-CoV RBD was not affected. Overall, mainchain H-bond interactions likely reduce sensitivity to RBD sidechain substitutions, making antibodies such as C118 and C022 more tolerant to differences between sarbecoviruses strains or variants.

### C118-S cryo-EM structure shows increased S trimer opening

On an S trimer, the class 4 cryptic epitope is at the base of the RBD, where it faces towards the center of the trimer (Barnes et al., 2020a; Huo et al., 2020; Yuan et al., 2020b). The epitope is buried in the closed, prefusion S conformation and interacts with portions of the spike S2 subunit and neighboring 'down' RBDs. Compared to class 2 or class 3 anti-RBD antibodies that recognize their epitopes in 'up' or 'down' RBD conformations (Barnes et al., 2020a), the class 4 epitope is less accessible and requires two 'up' RBDs for antibody binding (Piccoli et al., 2020a). Additionally, class 4 antibody binding may also require RBD rotation to prevent steric clashes with neighboring 'up' RBDs, as observed for the complexes of S trimer with EY6A, S2A4, and S304 (Piccoli et al., 2020a; Zhou et al., 2020a).

Given the similar binding poses of C118 and C022 antibodies, which bind with a more acute angle with respect to the RBD than EY6A or CR3022 (Figure 2E), and the increased breadth and potency of C118 and C022 relative to other class 4 anti-RBD antibodies (Figure 1E), we sought to understand the requirements for epitope recognition on a S trimer. Thus, we solved single-particle cryo-EM structures of C118 Fabs bound to SARS-CoV-2 S 6P trimers (Hsieh et al., 2020), finding two distinct states defined by RBDs adopting various rotational conformations (Figure 5A,B and Figure S6) as well as C118 Fab bound to dissociated S1 subunit protomers (Figure S6B). For the state 1 C118-S trimer complex structure solved to 3.4Å, we subsequently used symmetry expansion and local refinement to generate a 3.7Å map of the C118 V<sub>H</sub>V<sub>L</sub> – RBD interface (Figure S6B-E).

The C118 pose with respect to the RBD observed in the C118 – SARS-CoV-2 S structure was similar to the C118 – SARS-CoV RBD crystal structure (Figure S6F), demonstrating consistent recognition of the cryptic epitope of both SARS-CoV-2 and SARS-CoV RBDs. Furthermore, the C118 binding pose was oriented higher on the RBD relative to other class 4 anti-RBD antibodies (Figure 5C), and was consistent with SPR competition data that suggested C118 would sterically hinder ACE2 binding to the same protomer (Figure S3).

Despite differences in binding poses relative to other class 4 antibodies (Figure 5C), C118 binding also resulted in RBD conformations displaced further from the trimer center relative to S2E12 (a class 1 anti-RBD neutralizing antibody) (Tortorici, 2020) and ACE2 (Yan et al., 2020) (Figure 5D). On average, class 4 anti-RBD antibody binding resulted in an ~15-20Å displacement of the RBD relative to ACE2-bound conformations, which likely results in destabilization of the spike trimer. Indeed, S1 shedding induced by class 4 antibodies has been described as a possible neutralization mechanism (Huo et al., 2020; Piccoli et al., 2020a; Wec et al., 2020). The presence of C118-S1 protomer classes in our cryo-EM data suggested that C118 also induces shedding (Figure S6), but the role S1 shedding and premature S-triggering plays in C118-mediated neutralization requires further investigation.

### **C118 and C022 neutralization of sarbecoviruses demonstrate differential effects of avidity enhancement**

Neutralization of SARS-CoV-2 and SARS-CoV by COVA1-16 was found to be mediated by avidity effects based on potent neutralization by the bivalent COVA1-16 IgG, but not the monovalent Fab (Liu et al., 2020a). To evaluate whether intra-spike crosslinking, one source of avidity enhancement for bivalent antibodies, was possible for C118 or C022 IgGs, we examined the C118-S trimer structure to ask whether the positioning of two Fabs on adjacent RBDs would be compatible with binding by a single IgG. As previously described for other anti-RBD IgGs, we compared the distance between residues near the C-termini of adjacent Fab C<sub>H</sub>1 domains to analogous distances in crystal structures of intact IgGs, setting a cut-off of  $\leq 65\text{\AA}$  as potentially allowing a single IgG to include both Fabs (Barnes et al., 2020a). The measured distance for the C-termini of adjacent Fab C<sub>H</sub>1 domains in the symmetric State 1 C118-S trimer structure was 41Å (Figure 5E), suggesting that intra-spike crosslinking would be possible for C118 IgGs bound to spike trimers. The asymmetric State 2 C118-S structure included distances of 50Å, 89Å, and 92Å (Figure 5E), also allowing intra-spoke crosslinking between one combination of two bound RBDs, as well as the potential for inter-spoke crosslinking between adjacent spikes on the virion surface. In comparison, no other class 4 anti-RBD Fab-S trimer structures showed measured distances that would be compatible with intra-spoke

crosslinking (Figure 5E), thus any potential avidity effects for those IgGs could only occur via inter-spike crosslinking.

To further evaluate whether avidity could also facilitate cross-reactive neutralization by the C118 and C022 antibodies, we compared neutralization of SARS-CoV-2, SARS-CoV, WIV1, and SHC014 by the bivalent C118 and C022 IgGs and by two monovalent forms of each antibody: a 50 kDa Fab and an IgG size-matched bispecific IgG containing only one relevant Fab. The bispecific IgGs included one C118 or C022 RBD-binding Fab and a second non-RBD-binding Fab derived from the HIV-1 antibody 3BNC117 (Scheid et al., 2011). To interpret neutralization results, we calculated molar neutralization ratios (MNRs) defined as:  $[\text{IC}_{50} \text{ Fab or bispecific IgG (nM)}/\text{IC}_{50} \text{ IgG (nM)}]$  (Klein and Bjorkman, 2010). In the absence of avidity effects resulting from either crosslinking within a spike trimer (intra-spike crosslinking) or cross-linking between adjacent spike trimers (inter-spike crosslinking), an MNR would be 2.0, which accounts for twice as many relevant Fabs in a bivalent IgG compared to its monovalent forms.

Using pseudotyped SARS-CoV-2, SARS-CoV, WIV1, and SHC014, we derived neutralization potencies of the bivalent IgG, monovalent bispecific IgG, and Fab forms of C118 and C022 and then calculated MNRs for the bivalent IgG to bispecific IgG comparison (bispecific MNR) and for the bivalent IgG to Fab comparison (Fab MNR) (Figure 5F). Comparisons between the Fab and bispecific IgG forms of monovalent antibody allowed evaluation of potential steric effects that could increase neutralization potencies for larger IgGs compared to smaller Fabs. With the exception of the low MNRs derived from the IgG comparison with the bispecific and Fab forms of C118 against SARS-CoV-2 (MNRs of 5 and 3, respectively), we found mostly high MNRs ranging from the lowest values of >11 and 28 for the MNRs for C118 against SARS-CoV (where 11 is a minimal estimate since the C118 bispecific was non-neutralizing) to the highest values of 708 and 1444 for the C022 bispecific and Fab MNRs against SHC014. Four of the bispecific to Fab MNR comparisons showed a two-fold or higher Fab MNR than the comparable bispecific MNR, suggesting that at least some of the increased potencies of the bivalent IgGs compared with their counterpart Fabs resulted from steric effects. However, six of the eight monovalent to bivalent comparisons exhibited MNRs well over 70, suggestive of strong avidity effects. By contrast, mean MNRs derived for broadly neutralizing anti-HIV-1 Env antibodies are  $\leq 10$  (Wang et al., 2017), consistent with the low spike density on HIV-1 virions that largely prevents inter-spike crosslinking, and the architecture of the HIV-1 Env trimer, which prohibits intra-spike crosslinking for all known HIV-1 broadly neutralizing antibodies (Klein and Bjorkman, 2010). Taken together with the analysis of the C118-S trimer structure, the observed avidity effects for C118 IgGs binding to WIV1 and SHC014 and for the related C022 IgGs binding to the four viruses tested could arise from intra-spike as well as inter-spike crosslinking.

The question as to why C118 exhibits little or no avidity effects for neutralization of SARS-CoV-2 and SARS-CoV is difficult to address since the same IgG showed strong avidity effects against WIV1 and SHC014, and C022, which binds similarly to C118, showed avidity effects in neutralization of all four pseudoviruses. These results could derive from different binding characteristics for C118 to the SARS-CoV-2 and SARS-CoV RBDs compared with C118 and C022 interactions with the other sarbecoviruses evaluated. Indeed, simulations of avidity effects demonstrated that some combinations of IgG concentration and antigen-binding affinity and kinetic constants showed no advantages of bivalent versus monovalent binding (Klein, 2009; Klein and Bjorkman, 2010). Thus

the effects of avidity are a complicated function of concentration and binding constants that preclude predictions in the absence of experimental data.

## Discussion

Concerns about coronaviruses having spillover potential as well as the increasing prevalence of SARS-CoV-2 variants necessitates identification of cross-reactive antibodies. Antibodies elicited against infectious viruses for which there are multiple circulating variants, either within an individual or the population, often show a trade-off between potency and breadth (Corti et al., 2010; Desrosiers et al., 2016). In the case of antibody responses against SARS-CoV-2, the cause of the current global pandemic, many strongly neutralizing antibodies have been isolated that block ACE2 receptor interactions (Barnes et al., 2020a; Dejnirattisai et al., 2021; Lee et al., 2021; Liu et al., 2020b; Piccoli et al., 2020b). However, the ACE2-binding region of the RBD also tends to accumulate amino acid changes, as evidenced by substitutions identified in the current SARS-CoV-2 variants of concern (Annavajhala et al., 2021; Faria et al., 2021; Rambaut et al., 2020; Tegally et al., 2020; Voloch et al., 2020; West et al., 2021; Zhang et al., 2021), thus reducing the potential efficacies of vaccines and monoclonal antibody therapies. Recent studies suggest that antibodies against the S2 subunit offer the potential of greater cross-reactivity across coronaviruses, but these antibodies generally lack strong neutralization potency (Sauer et al., 2021; Shah et al., 2021; Wang et al., 2021).

The class 4 RBD-binding epitope, which is more conserved than the class 1 and class 2 RBD epitopes, represents a plausible target for the elicitation of antibodies with broad cross-reactive recognition across sarbecoviruses. Indeed, some recently described class 4 antibodies (e.g., CR3022, H014, COVA1-16, EY6A, ADI-56046) neutralize two or more sarbecovirus strains, and/or can bind RBDs from multiple sarbecoviruses (Liu et al., 2020a). However, while many class 4 antibodies show some cross-reactivity, they generally exhibit decreased potencies against heterologous sarbecovirus strains. For example, the SARS-CoV-derived CR3022 antibody does not potently neutralize SARS-CoV-2 (Huo et al., 2020), and the SARS-CoV-2 – derived COVA1-16 antibody does not potently neutralize SARS-CoV or SARS-CoV-2 variants of concern (Liu et al., 2020a; Liu et al., 2021).

Here we characterized two antibodies, C118 and C022, derived from different COVID-19 convalescent donors (Robbiani et al., 2020), which show breadth of and potent neutralization against sarbecoviruses of all three clades. The structural similarity of RBD binding poses between C022 and COVA1-16 (Liu et al., 2020a), which was derived from yet a third COVID-19 convalescent donor (Brouwer et al., 2020), suggests that these sorts of cross-reactive antibodies are commonly elicited by natural infection and that their epitope represents an attractive target for immunogen design. Of particular importance for the current pandemic, circulating variants of concern or variants of interest did not confer resistance to the C118 and C022 antibodies. In addition, C118 and C022 antibodies were not affected by naturally-occurring RBD mutations that undermine the activity of several antibodies approved for therapeutic use (Hoffmann et al., 2021; Starr et al., 2021b).

Analysis of our C118-RBD and C022-RBD complex structures revealed key details of cross-reactive recognition and broad sarbecovirus neutralization. First, C118 and C022 utilize long CDRH3s to facilitate interactions with the cryptic RBD epitope at the base of the RBD. In contrast to less potent

class 4 antibodies such as CR3022 (Huo et al., 2020; Yuan et al., 2020a; Yuan et al., 2020b) and EY6A (Zhou et al., 2020a) that also contact this region, the longer CDRH3 provides the opportunity to target a highly-conserved patch of residues across sarbecoviruses with an orientation that extends the epitope upwards to the ACE2 binding site. This extended epitope and many of the structural features of C118 and C022 binding are shared with COVA1-16 (Liu et al., 2020a). Second, the aforementioned binding poses of C118, C022, and COVA1-16, as well as overlap of the C022 epitope with the edge of the ACE2 binding site, suggested competition with ACE2 as part of their neutralization mechanisms. Indeed, competition experiments reported here for C118 and C022 and by others for COVA1-16 (Liu et al., 2020a) demonstrated competition with ACE2 for SARS-CoV-2 RBD binding. Third, C118, C022, and COVA1-16 formed many interactions with backbone atoms of RBD residues, adding a second level of buffering against viral escape since amino acid substitutions at these positions are less likely to abrogate antibody binding. Finally, the demonstration that C118 and C022 bivalence increased potency of neutralization against some of the viruses evaluated showed the potential for these antibodies to utilize avidity effects for neutralization of sarbecoviruses. Given the requirement for two 'up' RBDs on a S trimer for class 4 antibody binding, bivalent binding within a single S trimer would be possible. Thus we suggest that intra-spoke cross-linking would be an advantage for neutralization of sarbecoviruses, where avidity effects likely play a role.

In conclusion, class 4 antibodies that access the cryptic RBD epitope and compete with ACE2 binding are important for understanding cross-reactivity of human SARS-CoV-2 antibody responses elicited by natural sarbecovirus infection. We suggest that potent class 4 anti-RBD antibodies could be used therapeutically to avoid resistance to SARS-CoV-2 variants of concern, perhaps after in vitro selection to further improve their potencies. Structural characterization of these antibodies could also be used to inform future immunogen design efforts to elicit cross-reactive antibodies against SARS-CoV-2 variants and other sarbecoviruses.

### Acknowledgements

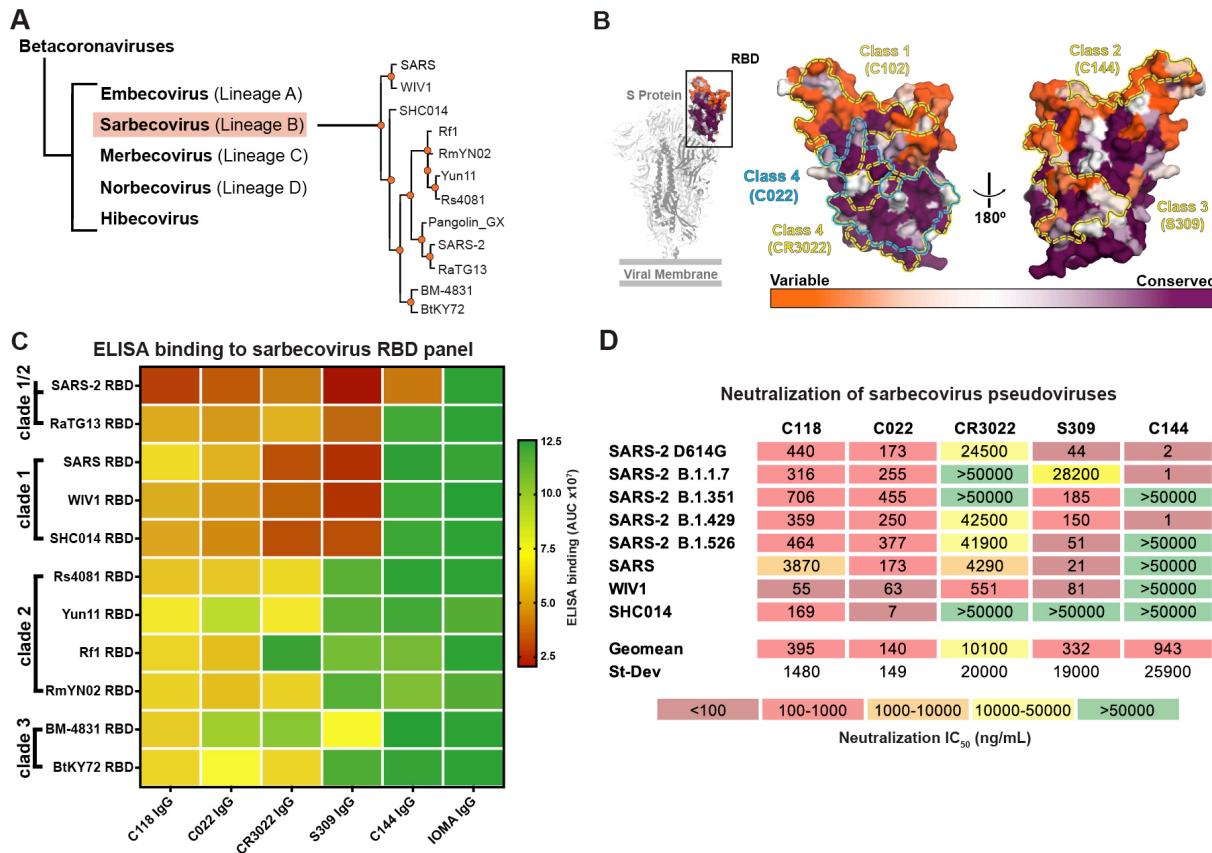
We thank J. Vielmetter, P. Hoffman, and the Protein Expression Center in the Beckman Institute at Caltech for expression assistance and K. Dam for assistance with soluble ACE2 purification. Electron microscopy was performed in the Caltech Cryo-EM Center with assistance from S. Chen and A. Malyutin. We thank the Gordon and Betty Moore and Beckman Foundations for gifts to Caltech to support the Molecular Observatory. We thank J. Kaiser, director of the Molecular Observatory at Caltech, and beamline staff C. Smith and S. Russi at SSRL for data collection assistance. Use of the Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, is supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No. DE-AC02-76SF00515. The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research, and by the National Institutes of Health, National Institute of General Medical Sciences (P30GM133894). The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of NIGMS or NIH. This work was supported by NIH (P01-AI138938-S1 to P.J.B. and M.C.N.), the Caltech Merkin Institute for Translational Research (P.J.B.), a George Mason University Fast Grant (P.J.B.), NIH R01AI078788 (to T.H.) and R01AI640511 (to P.D.B.). C.O.B was supported by the Hanna Gray Fellowship Program from the Howard Hughes Medical Institute and the Postdoctoral Enrichment Program from the Burroughs Wellcome Fund. M.C.N. is an HHMI investigator.

**Author Contributions**

C.A.J., A.A.C., P.J.B., and C.O.B. conceived and designed experiments. Proteins were produced and characterized by A.A.C., K.H.T., C.O.B., and C.A.J. Binding and neutralization studies were done by A.A.C., and F.M. with assistance from P.N.P.G., Y.L., and F.S. SPR binding competition experiments were done by C.A.J. with assistance from J.R.K. Structural studies were performed by C.A.J. with assistance from C.O.B. Structure analysis was done by C.A.J. with assistance from C.O.B. and A.A.C. Sequence analysis was done by A.P.W. Paper was written by C.A.J., P.J.B., and C.O.B. with assistance from A.A.C., T.H., M.C.N., P.D.B., and other authors.

**Declaration of interest**

The Rockefeller University has filed provisional patent applications in connection with this work on which M.C.N. (US patent 63/021,387) is listed as an inventor.

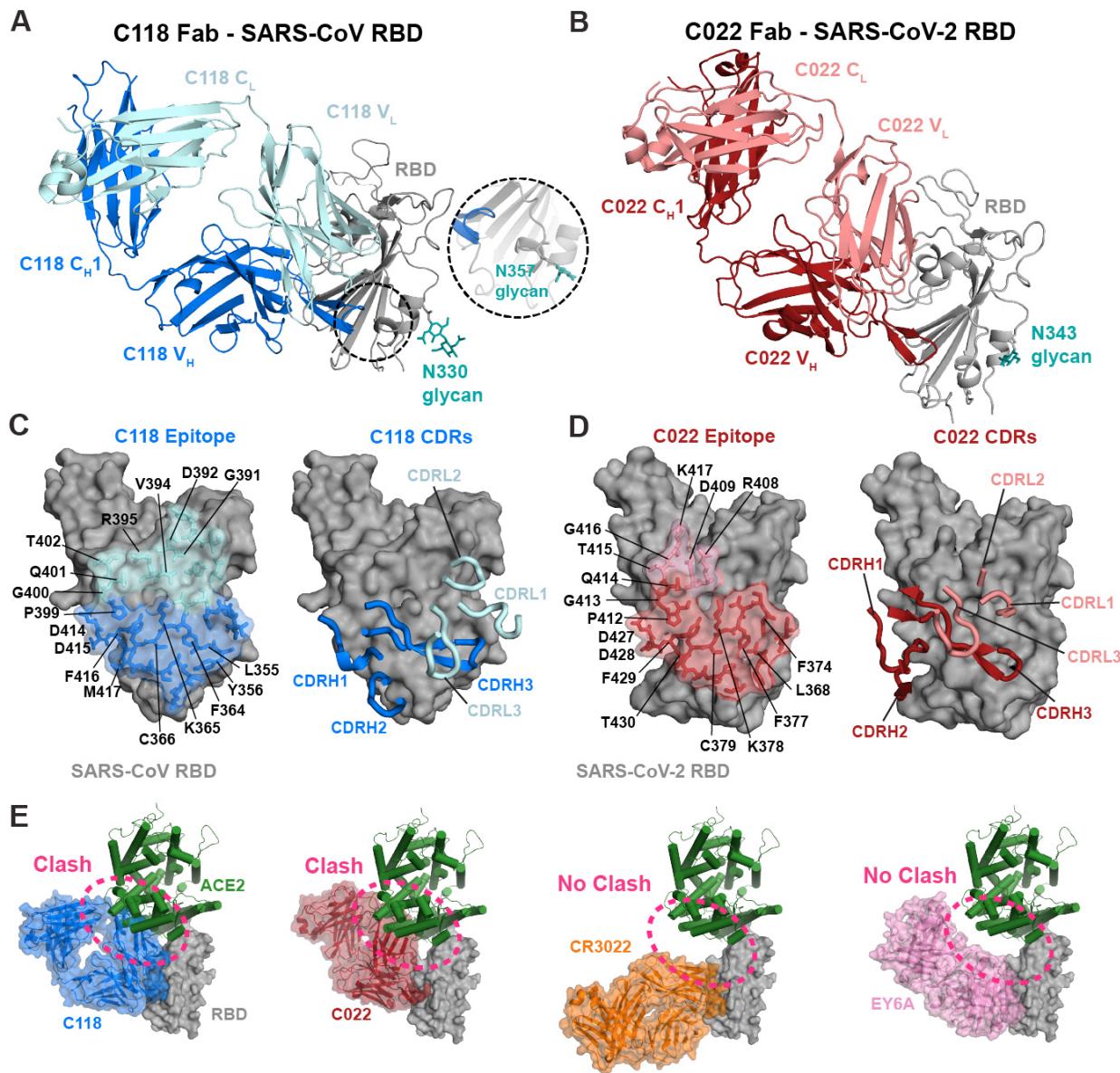


**Figure 1. C118 and C022 show diverse binding and neutralization of sarbecoviruses**

(A) Sarbecovirus (Lineage B) phylogenetic tree classified based on RBD sequence conservation. (B) Left: Cartoon rendering of SARS-CoV-2 S trimer (PDB 6VYB) showing location of 'up' RBD (surface, orange and purple). Right: Amino acid sequence conservation of 12 RBDs calculated as described (Landau et al., 2005) plotted on a surface representation of a SARS-CoV-2 RBD structure (PDB 7BZ5). Primary RBD epitopes for the indicated representatives from defined classes of RBD-binding antibodies (class 1-4) (Barnes et al., 2020a) are indicated as yellow dotted lines (PDB 7K90, 6W41, 7JX3,7K8M) . C022 epitope indicated as blue dotted line.

(C) Comparison of binding of the indicated monoclonal IgGs to a panel of sarbecovirus RBDs from ELISA data shown as area under the curve (AUC) values. Data presented are mean AUC values from two independent experiments. IOmA IgG is an anti-HIV-1 antibody serving as a negative control (Gristick et al., 2016).

(D) Neutralization IC<sub>50</sub> values for the indicated IgGs against SARS-CoV-2 (D614G version of the original variant (GenBank: NC\_045512)), SARS-CoV-2 variants of concern, and other ACE2-tropic sarbecovirus pseudoviruses. Geomean = geometric mean IC<sub>50</sub> in which IC<sub>50</sub> values >50000ng/mL were entered as 50000 ng/mL for the calculation. SD = standard deviation. IC<sub>50</sub> values are means of 2-7 independent experiments.

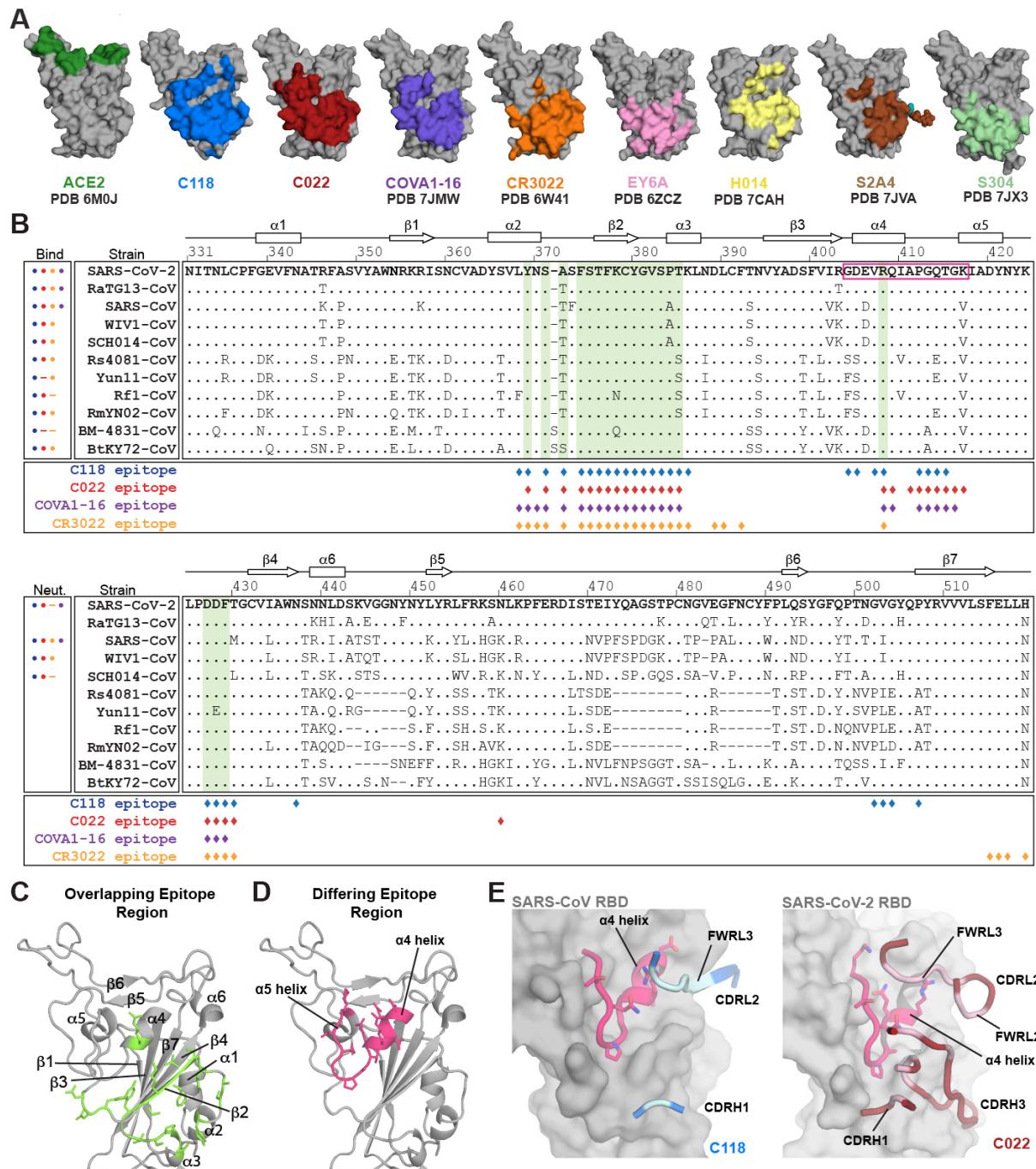


**Figure 2. Crystal structures of C022 and C118 Fabs bound to RBDs reveal class 4-like RBD binding.**

(A,B) Cartoon renderings of crystal structures of (A) C0118 Fab complexed with SARS-CoV RBD, and (B) C022 Fab complexed with SARS-CoV-2 RBD. Dashed circle shows location of SARS-CoV N357<sub>RBD</sub> residue, with the inset showing the N357<sub>RBD</sub> asparagine and glycan modeled based on the SARS-CoV spike-S230 structure (PDB 6NB6).

(C,D) CDR loops and RBD epitope residues of (C) C118 Fab and (D) C022 Fab overlaid on RBDs represented as gray surfaces with stick representations of epitope residues. Framework region residues, which account for some of the contacts for both antibodies, are not shown in right panels.

(E) Comparison of Fab poses for binding to an RBD-ACE2 complex. C118 Fab (blue), C022 Fab (red), CR3022 Fab (PDB 6W41; orange), and EY6A Fab (PDB 6CZC pink) modeled onto an ACE2-RBD structure (PDB 6M0J; RBD shown as a gray surface and ACE2 shown as a green cartoon).



**Figure 3.** The C118 and C022 epitopes include a conserved RBD helix.

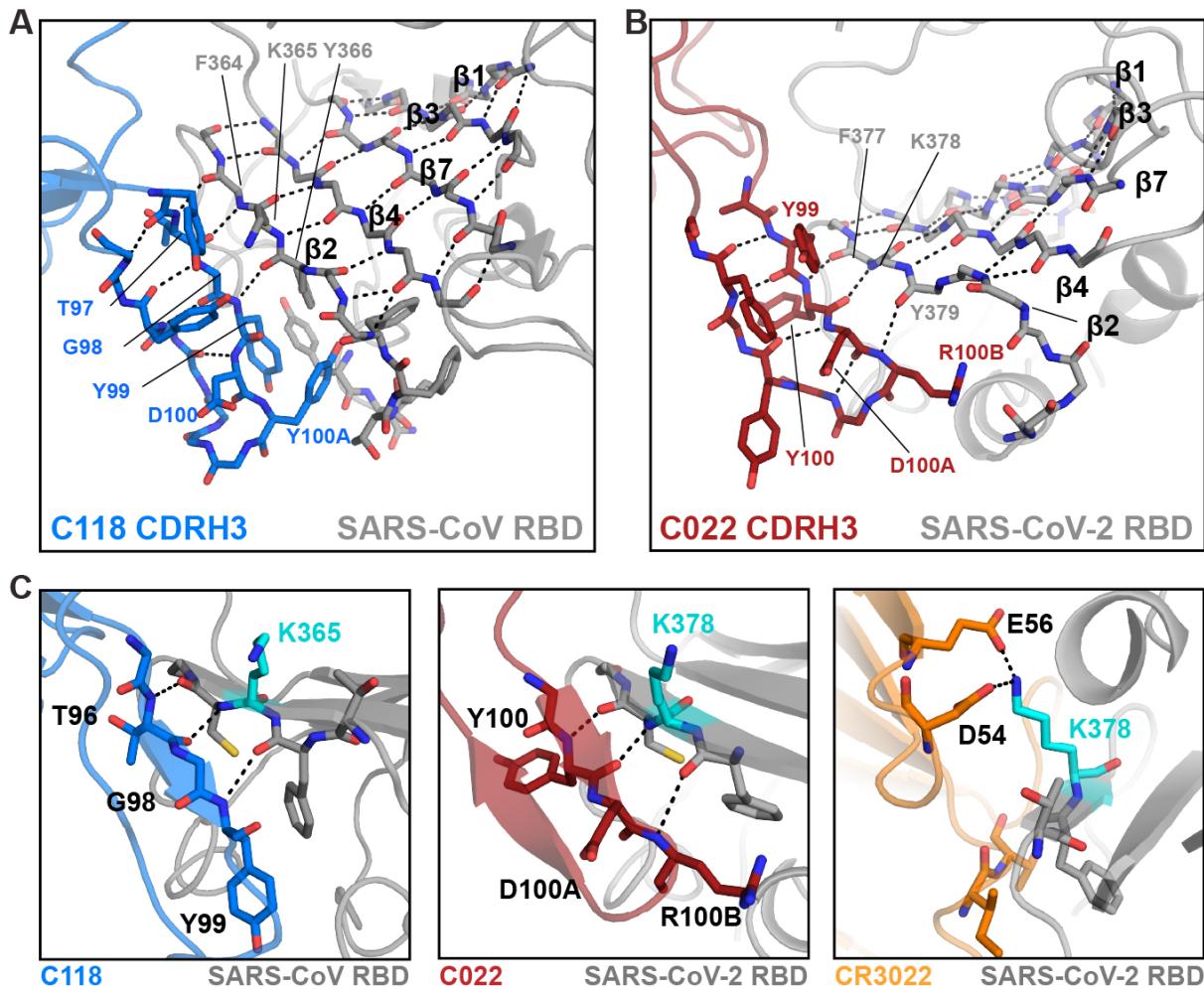
(A) Epitopes for ACE2 and monoclonal antibodies calculated from analyses of crystal structures of RBD complexes (human antibodies isolated from COVID-19 patients are C118, C022, COVA1-16, EY6A, and S2A4). RBDs shown are derived from SARS-CoV-2 except for the C118 panel, which is SARS-CoV RBD.

(B) Alignment of sequences for sarbecovirus RBDs (residue numbering for SARS-CoV-2 RBD). Secondary structure for SARS-CoV-2 RBD shown above alignment. Dots designate binding or neutralization for C118 (blue), C022 (red), or CR3022 (orange) for each strain. Diamonds designate RBD epitope residues for C118 binding to SARS-CoV (blue) and C022 (red) or CR3022 (orange) binding to SARS-CoV-2. Left boxes show binding by ELISA or neutralization of pseudovirus for each antibody for each strain; data for COVA1-16 from (Liu et al., 2020a). Circles show binding or neutralization, blank spaces designate not tested, and dashes designate no binding or neutralization. Shadings in the sequence alignment indicate conserved portions of epitope (green). Colored boxes show differing portion of epitope covering the  $\alpha 4$  helix and following loop (pink).

(C) Cartoon representation of SARS-CoV-2 RBD (gray) showing overlapping antibody-interacting residues (green) as sticks in epitopes for C118, C022, COVA1-16, and CR3022 (corresponding to green shading in panel B).

(D) Cartoon representation of SARS-CoV-2 RBD (gray) showing  $\alpha 4$  helix and following (sticks, pink) that differ in their contacts with C118, C022, COVA1-16, and CR3022 (pink shading in panel B).

(E) Cartoon representation of RBDs showing  $\alpha 4$  region of RBD and C118 (left) or C022 (right) interacting loops with interacting Fab residues in light blue (C118) and light pink (C022).

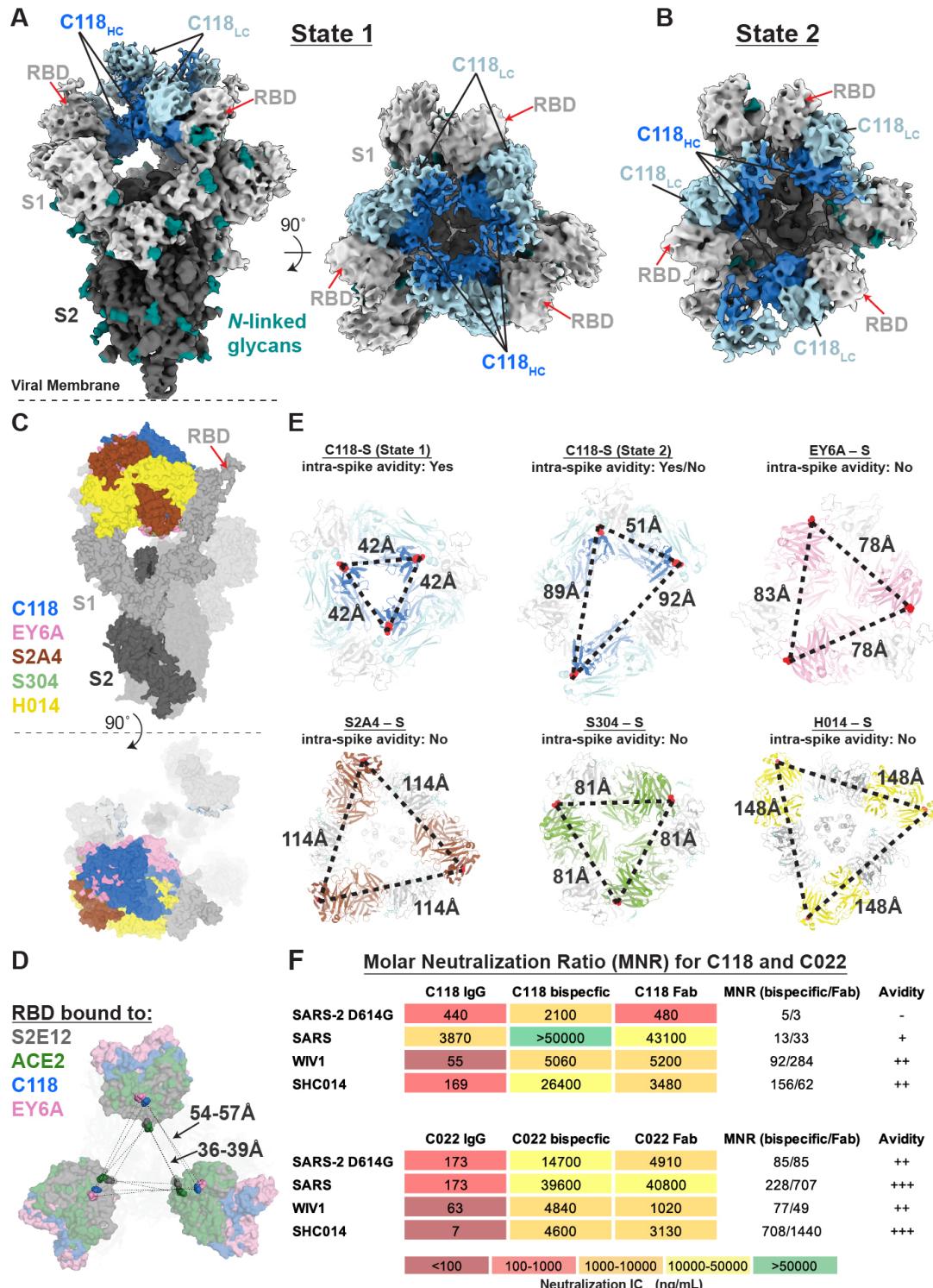


**Figure 4: C118 and C022 Fabs primarily use their CDRH3s for contacts with RBD alternative binding contacts for  $K378_{RBD}$ .**

(A) Close-up cartoon showing  $\beta$ -hairpin formed by C118 CDRH3 (blue sticks) and  $\beta$ -sheet formation with SARS-CoV RBD (grey cartoon with sticks). H-bonds shown as black dashed lines.

(B) Close-up cartoon showing  $\beta$ -hairpin formed by C022 CDRH3 (red sticks) and  $\beta$ -sheet formation with SARS-CoV-2 RBD (grey cartoon with sticks). H-bonds shown as black dashed lines.

(C) Cartoon and stick representation of C118-RBD (left), C022-RBD (middle) and CR3022-RBD (right) showing distinct interactions with residue  $K365_{SARS}/K378_{SARS2}$  (cyan).



**Figure 5.** Cryo-EM structure of C118-S complex shows binding to cryptic epitope in an 'up' RBD con-

**formation.**

(A) 3.4Å cryo-EM density for the C118 – S trimer complex (State 1). Side view (left panel) illustrates orientation with respect to the viral membrane (dashed line). Top view (right panel) shows symmetric binding at the trimer apex with C118 HC (blue) oriented in the interior.

(B) 4.4Å cryo-EM density for the C118 – S trimer complex (State 2). Top view illustrates asymmetry of complex due to RBD rotation in one protomer.

(C) Composite model of an open SARS-CoV-2 trimer bound by class 4 Fabs: C118 (this paper, blue), EY6A (PDB 6ZDH, pink), S2A4 (PDB 7JVC, brown), the class 4 anti-SARS antibody S304 (PDB 7JW0, green), and H014 (PDB 7CAK, yellow).

(D) Comparison of S trimer openness by measurements of C $\alpha$  distances for D428<sub>RBD</sub> between adjacent 'up' RBDs in S trimers complexed with: the class 1 antibody S2E12 (PDB 7K43, gray), soluble ACE2 (PDB 7KMS, green) and the class 4 antibodies C118 (this study, blue) and EY6A (PDB 6ZDH, pink).

(E) Prediction of potential intra-spike avidity effects by measurement of C $\alpha$  distances between the C-termini of adjacent C<sub>H</sub>1 domains for the mAb-S trimer complexes described in panel C. Measurements were used to evaluate the potential for intra-spike crosslinking by an IgG binding to a single spike trimer as described (Barnes et al., 2020a). For the H014-S complex, the C<sub>H</sub>1-C<sub>L</sub> domains were rigid body fit into the cryo-EM density (EMD-30333) prior to measurements.

(F) IC<sub>50</sub> values and molar neutralization ratios (MNRs) defined as: [IC<sub>50</sub> Fab or bispecific IgG (nM)/IC<sub>50</sub> IgG (nM)] (Klein and Bjorkman, 2010) for C118 and C022. IC<sub>50</sub> values shown for the IgGs are from Figure 1D. IC<sub>50</sub> values for all assays against SARS-CoV-2 and SARS-CoV are means of 2-7 independent experiments. Two MNRs are presented in the MNRs (bispecific/Fab) column: the MNR calculated using a bispecific IgG versus the bivalent IgG (left) and the MNR calculated using a Fab versus the bivalent IgG (right). Neutralization results with MNRs  $\leq$  5 are indicated as not demonstrating avidity effects (-),  $>$  10 are indicated as demonstrating minimal avidity (+), results with one MNR  $>$  50 are indicated as moderate avidity (++) and MNRs demonstrating strong avidity effects (one MNR  $>$  700) are indicated as (+++).

## References

Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., *et al.* (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66, 213-221.

Annajhala, M.K., Mohri, H., Zucker, J.E., Sheng, Z., Wang, P., Gomez-Simmonds, A., Ho, D.D., and Uhlemann, A.-C. (2021). A Novel SARS-CoV-2 Variant of Concern, B.1.526, Identified in New York. *medRxiv* 10.1101/2021.02.23.21252259.

Barnes, C.O., Jette, C.A., Abernathy, M.E., Dam, K.-M.A., Esswein, S.R., Gristick, H.B., Malyutin, A.G., Sharaf, N.G., Huey-Tubman, K.E., Lee, Y.E., *et al.* (2020a). SARS-CoV-2 neutralizing antibody structures inform therapeutic strategies. *Nature* 588, 682-687.

Barnes, C.O., West, A.P., Jr., Huey-Tubman, K.E., Hoffmann, M.A.G., Sharaf, N.G., Hoffman, P.R., Koranda, N., Gristick, H.B., Gaebler, C., Muecksch, F., *et al.* (2020b). Structures of Human Antibodies Bound to SARS-CoV-2 Spike Reveal Common Epitopes and Recurrent Features of Antibodies. *Cell* 182, 828-842 e816.

Bell, J.M., Chen, M., Baldwin, P.R., and Ludtke, S.J. (2016). High resolution single particle refinement in EMAN2.1. *Methods* 100, 25-34.

Brouwer, P.J.M., Caniels, T.G., van der Straten, K., Snitselaar, J.L., Aldon, Y., Bangaru, S., Torres, J.L., Okba, N.M.A., Claireaux, M., Kerster, G., *et al.* (2020). Potent neutralizing antibodies from COVID-19 patients define multiple targets of vulnerability. *Science* 369, 643-650.

Bunkóczki, G., and Read, R.J. (2011). Improvement of molecular-replacement models with Sculptor. *Acta Cryst* 67, 303-312.

Cao, Y., Su, B., Guo, X., Sun, W., Deng, Y., Bao, L., Zhu, Q., Zhang, X., Zheng, Y., Geng, C., *et al.* (2020). Potent neutralizing antibodies against SARS-CoV-2 identified by high-throughput single-cell sequencing of convalescent patients' B cells. *Cell* 10.1016/j.cell.2020.05.025.

Chan, K.K., Dorosky, D., Sharma, P., Abbasi, S.A., Dye, J.M., Kranz, D.M., Herbert, A.S., and Procko, E. (2020). Engineering human ACE2 to optimize binding to the spike protein of SARS coronavirus 2. *Science* 369, 1261-1265.

Chen, V.B., Arendall, W.B., 3rd, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* 66, 12-21.

Cohen, A.A., Gnanapragasam, P.N.P., Lee, Y.E., Hoffman, P.R., Ou, S., Kakutani, L.M., Keeffe, J.R., Wu, H.-J., Howarth, M., West, A.P., *et al.* (2021). Mosaic nanoparticles elicit cross-reactive immune responses to zoonotic coronaviruses in mice. *Science* 371, 735-741.

Corti, D., Sugitan, A.L., Pinna, D., Silacci, C., Fernandez-Rodriguez, B.M., Vanzetta, F., Santos, C., Luke, C.J., Torres-Velez, F.J., Temperton, N.J., *et al.* (2010). Heterosubtypic neutralizing antibodies are produced by individuals immunized with a seasonal influenza vaccine. *Journal of Clinical Investigation* 120, 1663-1673.

Crawford, K.H.D., Eguia, R., Dingens, A.S., Loes, A.N., Malone, K.D., Wolf, C.R., Chu, H.Y., Tortorici, M.A., Veesler, D., Murphy, M., *et al.* (2020). Protocol and Reagents for Pseudotyping Lentiviral Particles with SARS-CoV-2 Spike Protein for Neutralization Assays. *Viruses* 12.

de Wit, E., van Doremale, N., Falzarano, D., and Munster, V.J. (2016). SARS and MERS: recent insights into emerging coronaviruses. *Nat Rev Microbiol* 14, 523-534.

Dejnirattisai, W., Zhou, D., Ginn, H.M., Duyvesteyn, H.M.E., Supasa, P., Case, J.B., Zhao, Y., Walter, T.S., Mentzer, A.J., Liu, C., *et al.* (2021). The antigenic anatomy of SARS-CoV-2 receptor binding domain. *Cell* 10.1016/j.cell.2021.02.032.

Desrosiers, R.C., Wagh, K., Bhattacharya, T., Williamson, C., Robles, A., Bayne, M., Garrity, J., Rist, M., Rademeyer, C., Yoon, H., *et al.* (2016). Optimal Combinations of Broadly Neutralizing Antibodies for Prevention and Treatment of HIV-1 Clade C Infection. *PLOS Pathogens* 12, e1005520.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nuc Acids Res* 32, 1792-1797.

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

Faria, N.R., Claro, I.M., Candido, D., Moyses Franco, L.A., Andrade, P.S., Coletti, T.M., Silva, C.A.M., Sales, F.C., Manuli, E.R., Aguilar, R.S., *et al.* (2021). Genomic characterisation of an emergent SARS-CoV-2 lineage in Manaus: preliminary findings. <https://virological.org/t/genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-manaus-preliminary-findings/586>.

Fung, T.S., and Liu, D.X. (2019). Human Coronavirus: Host-Pathogen Interaction. *Annu Rev Microbiol* 73, 529-557.

Goddard, T.D., Huang, C.C., and Ferrin, T.E. (2007). Visualizing density maps with UCSF Chimera. *J Struct Biol* 157, 281-287.

Goddard, T.D., Huang, C.C., Meng, E.C., Pettersen, E.F., Couch, G.S., Morris, J.H., and Ferrin, T.E. (2018). UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci* 27, 14-25.

Greaney, A.J., Loes, A.N., Crawford, K.H.D., Starr, T.N., Malone, K.D., Chu, H.Y., and Bloom, J.D. (2021). Comprehensive mapping of mutations in the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human plasma antibodies. *Cell Host & Microbe* 29, 463-476.e466.

Gristick, H.B., von Boehmer, L., West, A.P., Jr., Schamber, M., Gazumyan, A., Goljanin, J., Seaman, M.S., Fatkenheuer, G., Klein, F., Nussenzweig, M.C., *et al.* (2016). Natively glycosylated HIV-1 Env structure reveals new mode for antibody recognition of the CD4-binding site. *Nat Struct Mol Biol* 23, 906-915.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 59, 307-321.

Haagmans, B.L., Al Dhahiry, S.H.S., Reusken, C.B.E.M., Raj, V.S., Galiano, M., Myers, R., Godeke, G.-J., Jonges, M., Farag, E., Diab, A., *et al.* (2014). Middle East respiratory syndrome coronavirus in dromedary camels: an outbreak investigation. *The Lancet Infectious Diseases* 14, 140-145.

Hoffmann, M., Arora, P., Groß, R., Seidel, A., Hörnich, B.F., Hahn, A.S., Krüger, N., Graichen, L., Hofmann-Winkler, H., Kempf, A., *et al.* (2021). SARS-CoV-2 variants B.1.351 and P.1 escape from neutralizing antibodies. *Cell* 10.1016/j.cell.2021.03.036.

Hoffmann, M., Kleine-Weber, H., Schroeder, S., Kruger, N., Herrler, T., Erichsen, S., Schiergens, T.S., Herrler, G., Wu, N.H., Nitsche, A., *et al.* (2020). SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* 181, 271-280 e278.

Hsieh, C.-L., Goldsmith, J.A., Schaub, J.M., DiVenere, A.M., Kuo, H.-C., Javanmardi, K., Le, K.C., Wrapp, D., Lee, A.G., Liu, Y., *et al.* (2020). Structure-based design of prefusion-stabilized SARS-CoV-2 spikes. *Science* 369, 1501-1505.

Huo, J., Zhao, Y., Ren, J., Zhou, D., Duyvesteyn, H.M.E., Ginn, H.M., Carrique, L., Malinauskas, T., Ruza, R.R., Shah, P.N.M., *et al.* (2020). Neutralization of SARS-CoV-2 by Destruction of the Prefusion Spike. *Cell Host & Microbe* 28, 445-454.e446.

Kabsch, W. (2010). XDS. *Acta Crystallogr D Biol Crystallogr* 66, 125-132.

Kirchdoerfer, R.N., Cottrell, C.A., Wang, N., Pallesen, J., Yassine, H.M., Turner, H.L., Corbett, K.S., Graham, B.S., McLellan, J.S., and Ward, A.B. (2016). Pre-fusion structure of a human coronavirus spike protein. *Nature* 531, 118-121.

Klein, J.S. (2009). Investigations in the design and characterization of HIV-1 neutralizing molecules (Pasadena: California Institute of Technology), pp. 166.

Klein, J.S., and Bjorkman, P.J. (2010). Few and far between: how HIV may be evading antibody avidity. *PLoS Pathog* 6, e1000908.

Korber, B., Fischer, W.M., Gnanakaran, S., Yoon, H., Theiler, J., Abfaltrerer, W., Hengartner, N., Giorgi, E.E., Bhattacharya, T., Foley, B., *et al.* (2020). Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. *Cell* 182, 812-827.e819.

Kreer, C., Zehner, M., Weber, T., Ercanoglu, M.S., Gieselmann, L., Rohde, C., Halwe, S., Korenkov, M., Schommers, P., Vanshylla, K., *et al.* (2020). Longitudinal Isolation of Potent Near-Germline SARS-CoV-2-Neutralizing Antibodies from COVID-19 Patients. *Cell* 10.1016/j.cell.2020.06.044.

Krissinel, E., and Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. *J Mol Biol* 372, 774-797.

Lan, J., Ge, J., Yu, J., Shan, S., Zhou, H., Fan, S., Zhang, Q., Shi, X., Wang, Q., Zhang, L., *et al.* (2020). Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* 581, 215-220.

Landau, M., Mayrose, I., Rosenberg, Y., Glaser, F., Martz, E., Pupko, T., and Ben-Tal, N. (2005). ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res* 33, W299-302.

Lee, B., Huang, K.-Y.A., Tan, T.K., Chen, T.-H., Huang, C.-G., Harvey, R., Hussain, S., Chen, C.-P., Harding, A., Gilbert-Jaramillo, J., *et al.* (2021). Breadth and function of antibody response to acute SARS-CoV-2 infection in humans. *PLoS Pathogens* 17, e1009352.

Lefranc, M.P., Giudicelli, V., Duroux, P., Jabado-Michaloud, J., Folch, G., Aouinti, S., Carillon, E., Duvergey, H., Houles, A., Paysan-Lafosse, T., *et al.* (2015). IMGT(R), the international ImMunoGeneTics information system(R) 25 years on. *Nucleic Acids Res* 43, D413-422.

Lefranc, M.P., Giudicelli, V., Ginestoux, C., Jabado-Michaloud, J., Folch, G., Bellahcene, F., Wu, Y., Gemrot, E., Brochet, X., Lane, J., *et al.* (2009). IMGT, the international ImMunoGeneTics information system. *Nucleic Acids Res* 37, D1006-1012.

Leist, S.R., Dinnon, K.H., Schäfer, A., Tse, L.V., Okuda, K., Hou, Y.J., West, A., Edwards, C.E., Sanders, W., Fritch, E.J., *et al.* (2020). A Mouse-Adapted SARS-CoV-2 Induces Acute Lung Injury and Mortality in Standard Laboratory Mice. *Cell* 183, 1070-1085.e1012.

Li, Q., Wu, J., Nie, J., Zhang, L., Hao, H., Liu, S., Zhao, C., Zhang, Q., Liu, H., Nie, L., *et al.* (2020). The Impact of Mutations in SARS-CoV-2 Spike on Viral Infectivity and Antigenicity. *Cell* 182, 1284-1294.e1289.

Li, W., Moore, M.J., Vasilieva, N., Sui, J., Wong, S.K., Berne, M.A., Somasundaran, M., Sullivan, J.L., Luzuriaga, K., Greenough, T.C., *et al.* (2003). Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* 426, 450-454.

Li, W., Shi, Z., Yu, M., Ren, W., Smith, C., Epstein, J.H., Wang, H., Cramer, G., Hu, Z., Zhang, H., *et al.* (2005). Bats are natural reservoirs of SARS-like coronaviruses. *Science* 310, 676-679.

Li, Z., Tomlinson, A.C., Wong, A.H., Zhou, D., Desforges, M., Talbot, P.J., Benlekbir, S., Rubinstein, J.L., and Rini, J.M. (2019). The human coronavirus HCoV-229E S-protein structure and receptor binding. *Elife* 8.

Liu, H., Wu, N.C., Yuan, M., Bangaru, S., Torres, J.L., Caniels, T.G., van Schooten, J., Zhu, X., Lee, C.-C.D., Brouwer, P.J.M., *et al.* (2020a). Cross-Neutralization of a SARS-CoV-2 Antibody to a Functionally Conserved Site Is Mediated by Avidity. *Immunity* 53, 1272-1280.e1275.

Liu, H., Yuan, M., Huang, D., Bangaru, S., Zhao, F., Lee, C.-C.D., Peng, L., Barman, S., Zhu, X., Nemazee, D., *et al.* (2021). A combination of cross-neutralizing antibodies synergizes to prevent SARS-CoV-2 and SARS-CoV pseudovirus infection. *Cell Host & Microbe* 10.1016/j.chom.2021.04.005.

Liu, L., Wang, P., Nair, M.S., Yu, J., Rapp, M., Wang, Q., Luo, Y., Chan, J.F.W., Sahi, V., Figueroa, A., *et al.* (2020b). Potent neutralizing antibodies against multiple epitopes on SARS-CoV-2 spike. *Nature* 584, 450-456.

Mastronarde, D.N. (2005). Automated electron microscope tomography using robust prediction of specimen movements. *J Struct Biol* 152, 36-51.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. *J Appl Crystallogr* 40, 658-674.

Muecksch, F., Weisblum, Y., Barnes, C.O., Schmidt, F., Schaefer-Babajew, D., Lorenzi, J.C.C., Flyak, A.I., DeLaitsch, A.T., Huey-Tubman, K.E., Hou, S., *et al.* (2021). Development of potency, breadth and resilience to viral escape mutations in SARS-CoV-2 neutralizing antibodies. *bioRxiv* 10.1101/2021.03.07.434227.

Piccoli, L., Park, Y.-J., Tortorici, M.A., Czudnochowski, N., Walls, A.C., Beltramello, M., Silacci-Fregni, C., Pinto, D., Rosen, L.E., Bowen, J.E., *et al.* (2020a). Mapping neutralizing and immunodominant sites on the SARS-CoV-2 spike receptor-binding domain by structure-guided high-resolution serology. *Cell* 10.1016/j.cell.2020.09.037.

Piccoli, L., Park, Y.-J., Tortorici, M.A., Czudnochowski, N., Walls, A.C., Beltramello, M., Silacci-Fregni, C., Pinto, D., Rosen, L.E., Bowen, J.E., *et al.* (2020b). Mapping Neutralizing and Immunodominant Sites on the SARS-CoV-2 Spike Receptor-Binding Domain by Structure-Guided High-Resolution Serology. *Cell* 183, 1024-1042.e1021.

Pinto, D., Park, Y.-J., Beltramello, M., Walls, A.C., Tortorici, M.A., Bianchi, S., Jaconi, S., Culap, K., Zatta, F., De Marco, A., *et al.* (2020). Cross-neutralization of SARS-CoV-2 by a human monoclonal SARS-CoV antibody. *Nature* 583, 290-295.

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

Rambaut, A., Pybus, O., Barclay, W., Barrett, J., Carabelli, A., Connor, T., Peacock, T., Robertson, D.L., Volz, E., and UK, C.-G.C. (2020). Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations. *virological.org*, <https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563>.

Rappazzo, C.G., Tse, L.V., Kaku, C.I., Wrapp, D., Sakharkar, M., Huang, D., Deveau, L.M., Yockachonis, T.J., Herbert, A.S., Battles, M.B., *et al.* (2021). Broad and potent activity against SARS-like viruses by an engineered human monoclonal antibody. *Science* 371, 823-829.

Robbiani, D.F., Gaebler, C., Muecksch, F., Lorenzi, J.C.C., Wang, Z., Cho, A., Agudelo, M., Barnes, C.O., Gazumyan, A., Finkin, S., *et al.* (2020). Convergent antibody responses to SARS-CoV-2 in convalescent individuals. *Nature* 584, 437-442.

Rogers, T.F., Zhao, F., Huang, D., Beutler, N., Burns, A., He, W.T., Limbo, O., Smith, C., Song, G., Woehl, J., *et al.* (2020). Rapid isolation of potent SARS-CoV-2 neutralizing antibodies and protection in a small animal model. *Science* 10.1126/science.abc7520.

Sauer, M.M., Tortorici, M.A., Park, Y.-J., Walls, A.C., Homad, L., Acton, O., Bowen, J., Wang, C., Xiong, X., de van der Schueren, W., *et al.* (2021). Structural basis for broad coronavirus neutralization. *bioRxiv* 10.1101/2020.12.29.424482.

Schaefer, W., Regula, J.T., Bahner, M., Schanzer, J., Croasdale, R., Durr, H., Gassner, C., Georges, G., Kettenberger, H., Imhof-Jung, S., *et al.* (2011). Immunoglobulin domain crossover as a generic approach for the production of bispecific IgG antibodies. *Proc Natl Acad Sci U S A* 108, 11187-11192.

Scharf, L., Wang, H., Gao, H., Chen, S., McDowall, A.W., and Bjorkman, P.J. (2015). Broadly Neutralizing Antibody 8ANC195 Recognizes Closed and Open States of HIV-1 Env. *Cell* 162, 1379-1390.

Scheid, J.F., Mouquet, H., Ueberheide, B., Diskin, R., Klein, F., Olivera, T.Y., Pietzsch, J., Fenyo, D., Abadir, A., Velinzon, K., *et al.* (2011). Sequence and Structural Convergence of Broad and Potent HIV Antibodies That Mimic CD4 Binding. *Science* 333, 1633-1637.

Schmidt, F., Weisblum, Y., Muecksch, F., Hoffmann, H.-H., Michailidis, E., Lorenzi, J.C.C., Mendoza, P., Rutkowska, M., Bednarski, E., Gaebler, C., *et al.* (2020). Measuring SARS-CoV-2 neutralizing antibody activity using pseudotyped and chimeric viruses. *Journal of Experimental Medicine* 217.

Schoofs, T., Barnes, C.O., Suh-Toma, N., Goljanin, J., Schommers, P., Gruell, H., West, A.P., Jr., Bach, F., Lee, Y.E., Nogueira, L., *et al.* (2019). Broad and Potent Neutralizing Antibodies Recognize the Silent Face of the HIV Envelope. *Immunity* 50, 1513-1529 e1519.

Schrödinger, L. (2011). The PyMOL Molecular Graphics System (The PyMOL Molecular Graphics System).

Seydoux, E., Homad, L.J., MacCamy, A.J., Parks, K.R., Hurlburt, N.K., Jennewein, M.F., Akins, N.R., Stuart, A.B., Wan, Y.H., Feng, J., *et al.* (2020). Analysis of a SARS-CoV-2-Infected Individual Reveals Development of Potent Neutralizing Antibodies with Limited Somatic Mutation. *Immunity* 53, 98-105 e105.

Shah, P., Canziani, G.A., Carter, E.P., and Chaiken, I. (2021). The Case for S2: The Potential Benefits of the S2 Subunit of the SARS-CoV-2 Spike Protein as an Immunogen in Fighting the COVID-19 Pandemic. *Frontiers in Immunology* 12.

Shi, R., Shan, C., Duan, X., Chen, Z., Liu, P., Song, J., Song, T., Bi, X., Han, C., Wu, L., *et al.* (2020). A human neutralizing antibody targets the receptor-binding site of SARS-CoV-2. *Nature* 584, 120-124.

Shu, Y., and McCauley, J. (2017). GISAID: Global initiative on sharing all influenza data - from vision to reality. *Euro Surveill* 22.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., *et al.* (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 7, 539.

Song, H.D., Tu, C.C., Zhang, G.W., Wang, S.Y., Zheng, K., Lei, L.C., Chen, Q.X., Gao, Y.W., Zhou, H.Q., Xiang, H., *et al.* (2005). Cross-host evolution of severe acute respiratory syndrome coronavirus in palm civet and human. *Proceedings of the National Academy of Sciences* 102, 2430-2435.

Starr, T.N., Czudnochowski, N., Zatta, F., Park, Y.-J., Liu, Z., Addetia, A., Pinto, D., Beltramello, M., Hernandez, P., Greaney, A.J., *et al.* (2021a). Antibodies to the SARS-CoV-2 receptor-binding domain that maximize breadth and resistance to viral escape. *bioRxiv* 10.1101/2021.04.06.438709.

Starr, T.N., Greaney, A.J., Addetia, A., Hannon, W.W., Choudhary, M.C., Dingens, A.S., Li, J.Z., and Bloom, J.D. (2021b). Prospective mapping of viral mutations that escape antibodies used to treat COVID-19. *Science* 371, 850-854.

Tegally, H., Wilkinson, E., Giovanetti, M., Iranzadeh, A., Fonseca, V., Giandhari, J., Doolabh, D., Pillay, S., San, E.J., Msomi, N., *et al.* (2020). Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa. *medRxiv* 10.1101/2020.12.21.20248640.

Terwilliger, T.C., Adams, P.D., Afonine, P.V., and Sobolev, O.V. (2018). A fully automatic method yielding initial models from high-resolution cryo-electron microscopy maps. *Nat Methods* 15, 905-908.

Tortorici, M.A. (2020). Ultrapotent human antibodies protect against SARS-CoV-2 challenge via multiple mechanisms. *Science* 370, 950-957.

Voloch, C.M., Silva F, R.d., de Almeida, L.G.P., Cardoso, C.C., Brustolini, O.J., Gerber, A.L., Guimarães, A.P.d.C., Mariani, D., Costa, R.M.d., Ferreira, O.C., *et al.* (2020). Genomic characterization of a novel SARS-CoV-2 lineage from Rio de Janeiro, Brazil. *medRxiv* 10.1101/2020.12.23.20248598.

Walls, A.C., Park, Y.J., Tortorici, M.A., Wall, A., McGuire, A.T., and Veesler, D. (2020). Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* 181, 281-292 e286.

Walls, A.C., Tortorici, M.A., Bosch, B.J., Frenz, B., Rottier, P.J.M., DiMaio, F., Rey, F.A., and Veesler, D. (2016). Cryo-electron microscopy structure of a coronavirus spike glycoprotein trimer. *Nature* 531, 114-117.

Wang, C., van Haperen, R., Gutiérrez-Álvarez, J., Li, W., Okba, N.M.A., Albulescu, I., Widjaja, I., van Dieren, B., Fernandez-Delgado, R., Sola, I., *et al.* (2021). A conserved immunogenic and vulnerable site on the coronavirus spike protein delineated by cross-reactive monoclonal antibodies. *Nature Communications* 12.

Wang, H., Gristick, H.B., Scharf, L., West, A.P., Galimidi, R.P., Seaman, M.S., Freund, N.T., Nussenzweig, M.C., and Bjorkman, P.J. (2017). Asymmetric recognition of HIV-1 Envelope trimer by V1V2 loop-targeting antibodies. *Elife* 6.

Wang, N., Li, S.-Y., Yang, X.-L., Huang, H.-M., Zhang, Y.-J., Guo, H., Luo, C.-M., Miller, M., Zhu, G., Chmura, A.A., *et al.* (2018). Serological Evidence of Bat SARS-Related Coronavirus Infection in Humans, China. *Virologica Sinica* 33, 104-107.

Wec, A.Z., Wrapp, D., Herbert, A.S., Maurer, D.P., Haslwanter, D., Sakharkar, M., Jangra, R.K., Dieterle, M.E., Lilov, A., Huang, D., *et al.* (2020). Broad neutralization of SARS-related viruses by human monoclonal antibodies. *Science* 369, 731-736.

Weisblum, Y., Schmidt, F., Zhang, F., DaSilva, J., Poston, D., Lorenzi, J.C.C., Muecksch, F., Rutkowska, M., Hoffmann, H.-H., Michailidis, E., *et al.* (2020). Escape from neutralizing antibodies by SARS-CoV-2 spike protein variants. *eLife* 9.

West, A.P., Barnes, C.O., Yang, Z., and Bjorkman, P.J. (2021). SARS-CoV-2 lineage B.1.526 emerging in the New York region detected by software utility created to query the spike mutational landscape. *bioRxiv* 10.1101/2021.02.14.431043;

West, A.P., Jr., Scharf, L., Horwitz, J., Klein, F., Nussenzweig, M.C., and Bjorkman, P.J. (2013). Computational analysis of anti-HIV-1 antibody neutralization panel data to identify potential functional epitope residues. *Proc Natl Acad Sci U S A* 110, 10598-10603.

Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G., McCoy, A., *et al.* (2011). Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* 67, 235-242.

Wrapp, D., Wang, N., Corbett, K.S., Goldsmith, J.A., Hsieh, C.L., Abiona, O., Graham, B.S., and McLellan, J.S. (2020). Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* 367, 1260-1263.

Yan, R., Zhang, Y., Li, Y., Xia, L., Guo, Y., and Zhou, Q. (2020). Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science* 367, 1444-1448.

Yuan, M., Liu, H., Wu, N.C., Lee, C.-C.D., Zhu, X., Zhao, F., Huang, D., Yu, W., Hua, Y., Tien, H., *et al.* (2020a). Structural basis of a shared antibody response to SARS-CoV-2. *Science* 10.1126/science.abd2321, eabd2321.

Yuan, M., Wu, N.C., Zhu, X., Lee, C.-C.D., So, R.T.Y., Lv, H., Mok, C.K.P., and Wilson, I.A. (2020b). A highly conserved cryptic epitope in the receptor binding domains of SARS-CoV-2 and SARS-CoV. *Science* 368, 630-633.

Yuan, Y., Cao, D., Zhang, Y., Ma, J., Qi, J., Wang, Q., Lu, G., Wu, Y., Yan, J., Shi, Y., *et al.* (2017). Cryo-EM structures of MERS-CoV and SARS-CoV spike glycoproteins reveal the dynamic receptor binding domains. *Nat Commun* 8, 15092.

Zhang, W., Davis, B.D., Chen, S.S., Sincuir Martinez, J.M., Plummer, J.T., and Vail, E. (2021). Emergence of a Novel SARS-CoV-2 Variant in Southern California. *Jama* 10.1001/jama.2021.1612.

Zhou, D., Duyvesteyn, H.M.E., Chen, C.-P., Huang, C.-G., Chen, T.-H., Shih, S.-R., Lin, Y.-C., Cheng, C.-Y., Cheng, S.-H., Huang, Y.-C., *et al.* (2020a). Structural basis for the neutralization of SARS-CoV-2 by an antibody from a convalescent patient. *Nature Structural & Molecular Biology* 27, 950-958.

Zhou, H., Ji, J., Chen, X., Bi, Y., Li, J., Hu, T., Song, H., Chen, Y., Cui, M., Zhang, Y., *et al.* (2021). Identification of novel bat coronaviruses sheds light on the evolutionary origins of SARS-CoV-2 and related viruses. *bioRxiv* 10.1101/2021.03.08.434390.

Zhou, P., Yang, X.L., Wang, X.G., Hu, B., Zhang, L., Zhang, W., Si, H.R., Zhu, Y., Li, B., Huang, C.L., *et al.* (2020b). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 579, 270-273.

Zivanov, J., Nakane, T., Forsberg, B.O., Kimanius, D., Hagen, W.J., Lindahl, E., and Scheres, S.H. (2018). New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* 7.

Zost, S.J., Gilchuk, P., Case, J.B., Binshtain, E., Chen, R.E., Nkolola, J.P., Schafer, A., Reidy, J.X., Trivette, A., Nargi, R.S., *et al.* (2020a). Potently neutralizing and protective human antibodies against SARS-CoV-2. *Nature* 584, 443-449.

Zost, S.J., Gilchuk, P., Chen, R.E., Case, J.B., Reidy, J.X., Trivette, A., Nargi, R.S., Sutton, R.E., Suryadevara, N., Chen, E.C., *et al.* (2020b). Rapid isolation and profiling of a diverse panel of human monoclonal antibodies targeting the SARS-CoV-2 spike protein. *Nat Med* 10.1038/s41591-020-0998-x.

## STAR Methods

### RESOURCE AVAILABILITY

#### Lead Contact

All requests for further information or reagents should be directed to the Lead Contact, Pamela Bjorkman (bjorkman@caltech.edu).

#### Materials Availability

All expression plasmids generated in this study for human CoV proteins, Fabs and IgGs are available upon request.

#### Data and Code Availability

Atomic models of C118 Fab complexed with SARS-CoV RBD and C022 Fab complexed with SARS-CoV-2 RBD have been deposited in the Protein Data Bank (PDB) (<http://www.rcsb.org/>) under accession codes **XXXX** and **YYYY**, respectively. The atomic model and cryo-EM map generated for the C118 Fab–SARS-CoV-2 S complex have been deposited at the PDB (<http://www.rcsb.org/>) and the Electron Microscopy Databank (EMDB) (<http://www.emdb-resource.org/>) under accession codes **AAAA** and **EMD-BBBB**, respectively.

### EXPERIMENTAL MODEL DETAILS

#### Cell lines

Cells for pseudovirus production (HEK293T) were cultured at 37°C and 5% CO<sub>2</sub> in Dulbecco's modified Eagle's medium (DMEM, Gibco) supplemented with 10% heat-inactivated fetal bovine serum (FBS, Sigma-Aldrich) and 5 µg/ml Gentamicin (Sigma-Aldrich).

Target cells for pseudovirus neutralization experiments (HEK293T<sub>ACE2</sub>) were generated as described (Robbiani et al., 2020) and cultured at 37°C and 5% CO<sub>2</sub> in Dulbecco's modified Eagle's medium (DMEM, Gibco) supplemented with 10% heat-inactivated fetal bovine serum (FBS, Sigma-Aldrich), 5 µg/ml gentamicin (Sigma-Aldrich), and 5 µg/mL Blasticidin (Gibco).

Expi293F cells (Gibco) for protein expression were maintained at 37°C and 8% CO<sub>2</sub> in Expi293 expression medium (Gibco), transfected using an Expi293 Expression System Kit (Gibco) and maintained under shaking at 130 rpm. All cell lines were female and were not specifically authenticated.

#### Bacteria

*E. coli* DH5 Alpha (Zymo Research) used for propagation of expression plasmids were cultured with shaking at 250 rpm at 37°C in LB broth (Sigma-Aldrich).

#### Viruses

To generate pseudotyped viral stocks, HEK293T cells were transfected with pNL4-3ΔEnv-nanoluc and pSARS-CoV2-S<sub>trunc</sub> (Robbiani et al., 2020) using polyethylenimine, leading to production of HIV-

1-based pseudovirions carrying the SARS-CoV-2 S protein at the surface. Eight hours after transfection, cells were washed twice with phosphate buffered saline (PBS) and fresh media was added. Supernatants containing pseudovirus were harvested 48 hours post transfection, filtered and stored at -80°C. Infectivity of pseudoviruses was determined by titration on 293T<sub>ACE2</sub> cells.

## METHOD DETAILS

### Phylogenetic trees

Sequence alignments of RBDs were made with Clustal Omega (Sievers et al., 2011). Phylogenetic trees were calculated from amino acid alignments using PhyML 3.0 (Guindon et al., 2010) and visualized with PRESTO (<http://www.atgc-montpellier.fr/presto>).

### Protein Expression

Fabs and IgGs were expressed and purified as previously described (Scharf et al., 2015; Schoofs et al., 2019) and stored at 4 °C. Bispecific IgGs (C118 or C022 plus 3BNC117, a non-coronavirus binding HIV-1 antibody (Scheid et al., 2011)) were produced by co-transfection of two heavy chain and two light chain genes that included knobs-into-holes mutations in IgG Fc and a domain cross-over in the 3BNC117 Fab to prevent incorrect light chain pairing (Schaefer et al., 2011). Antibody CDR lengths were determined using the IMGT definitions (Lefranc et al., 2015; Lefranc et al., 2009).

The following C-terminally 6xHis-tagged RBD proteins were transfected and expressed as described previously (Cohen et al., 2021): SARS-CoV-2 RBD (residues 328-533), SARS-CoV-2 RBD mutants (residues 319-541), SARS RBD (residues 318-510), SHC014 RBD (residues 307-524), WIV-1 RBD (residues 307-528), RaTG13 RBD (residues 319-541), Rs4081 RBD (residues 310-515), Yun11 RBD (residues 310-515), Rf1 RBD (residues 310-515), RmYN02 RBD (298-503), BM-4831 RBD (residues 310-530), BtKY72 RBD (residues 309-530). A trimeric SARS-CoV-2 ectodomain (residues 16-1206 of the early SARS-CoV-2 GenBank MN985325.1 sequence isolate with 6P (Hsieh et al., 2020) stabilizing mutations, a mutated furin cleavage site between S1 and S2, a C-terminal TEV site, foldon trimerization motif, octa-His tag, and AviTag) was expressed as described (Barnes et al., 2020a; Barnes et al., 2020b). A gene encoding a 6xHis-tagged soluble human ACE2 construct (residues 1-615) was purchased from Addgene (Catalog # 149268) and expressed and purified as described (Chan et al., 2020).

SARS-CoV-2 S trimer, RBDs, and soluble ACE2 were purified by Nickel-NTA and size-exclusion chromatography using a Superdex 200 column (GE Life Sciences) as described (Barnes et al., 2020a; Cohen et al., 2021). Peak fractions were identified by SDS-PAGE, and those containing S trimer, monomeric RBDs, or soluble ACE2 were pooled, concentrated, and stored at 4 °C (RBDs) or flash frozen in nitrogen and stored at -80 °C (S trimer) until use.

### ELISAs

Purified RBD at 10 µg/ml in 0.1 M NaHCO<sub>3</sub> pH 9.8 was coated onto Nunc® MaxiSorp™ 384-well plates (Sigma) and stored overnight at 4°C. The following day, plates were blocked with 3% bovine serum albumin (BSA) in TBS-T Buffer (TBS + 0.1% Tween20) for 1hr at room temperature. Blocking solution was removed from the plates, purified IgGs at 50 µg/mL were serially diluted by 4-fold with TBS-T/3% BSA and added to plates for 3 hr at room temperature. Plates were washed with

TBS-T and then incubated with 1:15,000 dilution of secondary HRP-conjugated goat anti-human IgG for 45 minutes at room temperature (Southern Biotech). Plates were washed again with TBS-T and developed using SuperSignal™ ELISA Femto Maximum Sensitivity Substrate (ThermoFisher) and read at 425 nm. ELISAs were done in duplicate, and curves were plotted and integrated to obtain the area under the curve (AUC) using Graphpad Prism v9.1.0.

### Neutralization assays

SARS-CoV-2, SARS-CoV-2 variants of concern (Annavajhala et al., 2021; Faria et al., 2021; Rambaut et al., 2020; Tegally et al., 2020; Voloch et al., 2020; West et al., 2021; Zhang et al., 2021), SARS-CoV, WIV1, and SHC014 pseudoviruses based on HIV-1 lentiviral particles were prepared as described (Cohen et al., 2021; Crawford et al., 2020; Robbiani et al., 2020) using genes encoding S protein sequences with cytoplasmic tail deletions: 21 amino acid deletions for SARS-CoV-2, SARS-CoV-2 variants of concern, WIV1, and SHC014 and a 19 amino acid deletion for SARS-CoV. Plasmids expressing the spike protein found in the bat (*Rinolophus Sinicus*) coronavirus bCoV-WIV16 as well as the pangolin (*Manis javanica*) coronaviruses from Guandong, China (pCoV-GD) and Guanxi, China (pCoV-GX) have been described previously and are based on ALK02457 (Genebank), Pangolin\_CoV\_EPI\_ISL\_410721(Gisaid) and Pangolin\_CoV\_EPI\_ISL\_410542 (Gisaid) (Muecksch et al., 2021).

Relative to the SARS-CoV-2 spike gene (Wuhan-Hu-1 Spike Glycoprotein Gene, D614G mutant, designated as 'wt' in Figure 1E), the SARS-CoV-2 variants of concern included the D614G mutation and the following other substitutions: B.1.351: L18F, D80A, D215G, del242-244, R246I, K417N, E484K, N501Y, A701V; B.1.1.7: del69-70, del144, N501Y, A570D, P681H, T716I, S982A, D1118H; B.1.429: S13I, W152C, L452R, and B.1.526: L5F, T95I, D253G, E484K, A701V. For neutralization assays presented in Figure 1E,F, four-fold dilutions of purified IgGs (starting concentrations of 50 µg/mL) were incubated with a pseudotyped virus for 1 hour at 37°C. Cells were washed twice with phosphate-buffered saline (PBS) and lysed with Luciferase Cell Culture Lysis 5x reagent (Promega) after incubation with 293T<sub>ACE2</sub> target cells for 48 hours at 37°C. NanoLuc Luciferase activity in lysates was measured using the Nano-Glo Luciferase Assay System (Promega). Relative luminescence units (RLUs) were normalized to values derived from cells infected with pseudotyped virus in the absence of IgG. Half-maximal inhibitory concentrations (IC<sub>50</sub> values) were determined using 4- or 5-parameter nonlinear regression in AntibodyDatabase (West et al., 2013).

Relative to the SARS-CoV-2 spike gene (Wuhan-Hu-1; NC\_045512, D614 sequence designated as 'wt' in Figure S2), a panel of plasmids expressing RBD mutant SARS-CoV-2 S proteins in the context of pSARS-CoV-2-S<sub>Δ19</sub> have been described previously (Muecksch et al., 2021; Robbiani et al., 2020; Schmidt et al., 2020; Weisblum et al., 2020). The E484K substitution was constructed in the context of a pSARS-CoV-2-S<sub>Δ19</sub> variant with a mutation in the furin cleavage site (R683G) to increase infectivity (Muecksch et al., 2021). The IC<sub>50</sub> values of this pseudotype (E484K/R683G) was compared to a wild-type SARS-CoV-2 S sequence carrying R683G in the subsequent analyses. For neutralization assays presented in Figure S2, monoclonal antibodies were four-fold serially diluted and incubated with SARS-CoV-2 pseudotyped HIV-1 reporter virus for 1 h at 37 °C (final starting concentration of 2.5 µg/ml). The antibody and pseudotyped virus mixture was added to HT1080ACE2.cl 14 cells (Schmidt et al., 2020). After 48 h, cells were washed with PBS and lysed with Luciferase Cell Culture Lysis 5× reagent (Promega). Nanoluc luciferase activity in cell lysates was measured using the Nano-Glo Luciferase Assay System (Promega) and the Glomax Navigator

(Promega). Relative luminescence units were normalized to those derived from cells infected with SARS-CoV-2 pseudovirus in the absence of monoclonal antibodies. The 50% inhibitory concentration ( $IC_{50}$ ) was determined using 4-parameter nonlinear regression (least-squares regression method without weighting; constraints: top = 1, bottom = 0) (GraphPad Prism).

### SPR-based ACE2 binding competition experiments

Surface Plasmon Resonance (SPR) experiments were done using a Biacore T200 instrument (GE Healthcare).

Surface Plasmon Resonance (SPR) experiments were done using a Biacore T200 instrument (GE Healthcare). Purified SARS CoV-2 RBD was conjugated to each of the four flow cells using primary amine chemistry at pH 4.5 (Biacore manual) to a CM5 chip (GE Healthcare) to a response level of ~700 resonance units (RUs). C118, C022, C144, and CR3022 IgG (1000nM) in buffer HBS-EP+ (150mM sodium chloride, 10mM HEPES, 3mM EDTA, 0.05% Tween-20, pH 7.6) were each injected on the RBD-CM5 chip for a contact time of 600 sec at 30 $\mu$ L/min. A second injection of soluble ACE2 at 250nM was injected over the immobilized RBD-Fab at 30 $\mu$ L/min for a contact time of 300 sec and dissociation time of 30 sec in HBS-EP+ buffer. Data were analyzed and plotted using Prism 9 (Graphpad).

### X-ray crystallography

Fab-RBD complexes were assembled by incubating an RBD with a 1.5x molar excess of Fab for 1 hr on ice followed by size exclusion chromatography on an S200 300/10 increase column (GE Life Sciences). Fractions containing complex were pooled and concentrated to 8mg/mL. Crystallization trials using commercially-available screens (Hampton Research) were performed at room temperature using the sitting drop vapor diffusion method by mixing equal volumes of a Fab-RBD complex and reservoir using a TTP LabTech Mosquito instrument. Crystals were obtained for C118 Fab-SARS RBD in 0.2M sodium fluoride, 20% w/v polyethylene glycol 3,350 and for C022 Fab-SARS-CoV-2 RBD in 0.05M ammonium sulfate, 0.05M Bis-Tris, 30% v/v pentaerythritol ethoxylate (15/4 EO/OH). Crystals were cryoprotected by adding glycerol directly to drops to a final concentration of 20% and then looped and cryopreserved in liquid nitrogen.

X-ray diffraction data were collected at the Stanford Synchrotron Radiation Lightsource (SSRL) beamline 12-2 on a Pilatus 6M pixel detector (Dectris). Data from single crystals were indexed and integrated in XDS (Kabsch, 2010) and merged using AIMLESS in CCP4 (Winn et al., 2011) (Table S1). The C022-RBD structure was solved by molecular replacement in PHASER (McCoy et al., 2007) using unmodified RBD coordinates (PDB 7K8M) and coordinates from C102 Fab (PDB 7K8M) after trimming heavy chain and light chain variable domains using Sculptor (Bunkóczki and Read, 2011) as search models. Coordinates were refined with *phenix.refine* from the PHENIX package ver. 1.17.1 (Adams et al., 2010) and cycles of manual building in Coot (ver 0.8.9.1) (Emsley et al., 2010) (Table S1).

### Cryo-EM Sample Preparation

C118 Fab-S trimer complex was assembled by incubating purified SARS-CoV-2 S trimer at a 1.2:1 molar excess of purified Fab per S protomer at RT for 30 min. 17  $\mu$ L of complex was mixed with 0.8 $\mu$ L of a 0.5% w/v F-octylmaltoside solution (Anatrace) and then 3 $\mu$ L were immediately applied to a 300 mesh, 1.2/1.3 AuUltraFoil grid (Electron Microscopy Sciences) that had been freshly glow

discharged for 1 min at 20mA using a PELCO easiGLOW (Ted Pella). The grid was blotted for 3.5s with Whatman No. 1 filter paper at 22°C and 100% humidity then vitrified in 100% liquid ethane using a Mark IV Vitrobot (FEI) and stored under liquid nitrogen.

### Cryo-EM data collection and processing

Single-particle cryo-EM data were collected for the C118-S trimer complex as previously described (Barnes et al., 2020a). Briefly, for the C118-S trimer complex, micrographs were collected on a Talos Arctica transmission electron microscope (Thermo Fisher) operating at 200 kV using a 3x3 beam image shift pattern with SerialEM automated data collection software (Mastronarde, 2005). Movies were obtained on a Gatan K3 Summit direct electron detector operating in counting mode at a nominal magnification of 45,000x (super-resolution 0.4345 Å/pixel) using a defocus range of -0.7 to -2.0 μm. Movies were collected with an 3.6 s exposure time with a rate of 13.5 e<sup>-</sup>/pix/s, which resulted in a total dose of ~60 e<sup>-</sup>/Å<sup>2</sup> over 40 frames. The 2,970 movies were patch motion corrected for beam-induced motion including dose-weighting within cryoSPARC v3.1 (Punjani et al., 2017) after binning super resolution movies by 2 (0.869 Å/pixel). The non-dose-weighted images were used to estimate CTF parameters using Patch CTF in cryoSPARC, and micrographs with poor CTF fits and signs of crystalline ice were discarded, leaving 2,487 micrographs. Particles were picked in a reference-free manner using Gaussian blob picker in cryoSPARC (Punjani et al., 2017). An initial 923,707 particle stack was extracted, binned x4 (3.48 Å/pixel), and subjected to *ab initio* volume generation (4 classes) and subsequent heterogeneous refinement. The 3D classes that showed features for a Fab-S trimer complex were 2D classified to identify class averages corresponding to intact S-trimer complexes with well-defined structural features. This routine resulted in a new particle stack of 110,789 particles, which were unbinned (0.836 Å/pixel) and re-extracted using a 432 box size. Particles were then moved to Relion v3.1 (Zivanov et al., 2018), for further 3D classification (k=6), which revealed two distinct states of the C118-S trimer complex.

Particles from state 1 (53,728 particles) and state 2 (31,422 particles) were separately refined using non-uniform 3D refinement imposing either C3 or C1 symmetry in cryoSPARC, respectively, to final resolutions of 3.4 Å and 4.5 Å according to the gold-standard FSC (Bell et al., 2016), respectively. To improve features at the C118-RBD interface, particles from State 1 were symmetry expanded and classified for a focused, non-uniform 3D local refinement in cryoSPARC. A soft mask was generated around the C118 V<sub>H</sub>V<sub>L</sub> – RBD domains (5-pixel extension, 10-pixel soft cosine edge) for local refinements. These efforts resulted in a modest improvement in the RBD-C118 Fab interface (Figure S6B), with an overall resolution of 3.7 Å according to the gold-standard FSC.

### Cryo-EM Structure Modeling and Refinement

Initial coordinates were generated by rigid-body docking reference structures into cryo-EM density using UCSF Chimera (Goddard et al., 2007). The following coordinates were used: SARS-CoV-2 S 6P trimer: PDB 7K4N (mutated to include 6P mutations), PDB 7BZ5, and C118 Fab variable domains: this study. These initial models were then refined into cryo-EM maps using one round of rigid body refinement, morphing and real space refinement in Phenix (Adams et al., 2010). Sequence-updated models were built manually in Coot (Emsley et al., 2010) and then refined using iterative rounds of refinement in Coot and Phenix (Adams et al., 2010). Glycans were modeled at possible N-linked glycosylation sites (PNGSs) in Coot using ‘blurred’ maps processed with a variety of B-factors (Terwilliger et al., 2018). Validation of model coordinates was performed using MolProbity (Chen et al., 2010) and is reported in Table S3.

### Structure Analyses

Interacting residues were determined using PDBePISA (Krissinel and Henrick, 2007) for the C118 and C022 epitopes using the following criteria: Potential H-bonds were assigned using a distance of  $<3.6\text{\AA}$  and an A-D-H angle of  $>90^\circ$ , and the maximum distance allowed for a van der Waals interaction was  $4.0\text{\AA}$ . H-bonds assigned for the C022-RBD complex should be considered tentative due to the relatively low resolution of the structure ( $3.2\text{\AA}$ ). Epitope patches for other antibodies in Figure 4A were defined as residues containing an atom within  $4\text{\AA}$  of the partner protein as determined in PyMOL (Schrödinger, 2011). Buried surface areas (BSAs) were determined with PDBePISA (Krissinel and Henrick, 2007) using a  $1.4\text{\AA}$  probe. Structure figures were made using PyMOL ver. 2.3.5 (Schrödinger, 2011) or UCSF Chimera ver. 1.14 (Goddard et al., 2018). Fab-RBD-ACE2 complex figures (Figure 2E) were made by aligning RBD  $\text{C}\alpha$  atoms of Fab-RBD (this study and PDBs 6W41 and 6ZCZ) and RBD-ACE2 structures (PDB 6M0J). As density at position  $\text{N}357_{\text{RBD}}$  for our C118-SARS RBD structure precluded building of the glycan, it was modeled (Figure 2A) by aligning  $\text{C}\alpha$  atoms of residues 353-371 of SARS-CoV spike-S230 structure (PDB 6NB6, chain E) and overlaying the glycan at  $\text{N}357_{\text{RBD}}$  from the SARS-CoV spike on the RBD model of the C118-RBD crystal structure. Sequence alignments were done using the MUSCLE server (<https://www.ebi.ac.uk/Tools/msa/muscle/>) (Edgar, 2004). Secondary structure was defined as described in Huo et al., 2020.

To predict whether intra-spike crosslinking by a single IgG binding to a spike trimer might be possible, we measured the distance between residue  $222_{\text{HC}}$   $\text{C}\alpha$  atoms in the  $\text{C}_{\text{H}1}$  domains of adjacent Fabs in Fab-S structures as previously described (Barnes et al., 2020a). This distance was compared to analogous distances in crystal structures of intact IgGs ( $42\text{\AA}$ , PDB 1HZH;  $48\text{\AA}$ , PDB 1IGY;  $52\text{\AA}$ , PDB 1IGT). We accounted for potential influences of crystal packing in intact IgG structures, flexibility in the  $\text{V}_{\text{H}}\text{-}\text{V}_{\text{L}}\text{/}\text{C}_{\text{H}1}\text{-}\text{C}_{\text{L}}$  elbow bend angle, and uncertainties in  $\text{C}_{\text{H}1}\text{-}\text{C}_{\text{L}}$  domain placement in Fab-S cryo-EM structures, by setting a cut-off of  $\leq 65\text{\AA}$  for this measured distance as potentially allowing for a single IgG to include both Fabs when binding a spike trimer.

**Supplemental Items**

**Table S1: Crystallographic data collection and validation statistics for C118-SARS RBD and C022-SARS2 RBD (related to Figure 2)**

**Table S2. Buried Surface Area (related to Figure 3)**

**Table S3. Cryo-EM data collection and refinement statistics for C118-S complex structure (related to Figure 4).**

**Supplemental Movie 1**

C118 and C022 Fab bind to SARS-CoV-2 RBD in a class 4 orientation at a 'cryptic' epitope. The video shows soluble SARS-CoV-2 spike (PDB: 7BZ5) and the 'down' and 'up' conformations of the RBD domain. Next, the sequence conservation of 12 sarbecovirus RBDs is mapped onto the surface of the RBD domain. Finally, The C118-SARS-CoV RBD and C022-SARS-CoV-2 RBD structures are shown.

**Table S1: Crystallographic data collection and validation statistics for C118-SARS RBD and C022-SARS2 RBD. (Related to Figure 2)**

PDB ID	C118 Fab – SARS-CoV RBD AAAA	C022 Fab – SARS-CoV-2 RBD BBBB
<b>Data collection<sup>a,b</sup></b>		
Space group	P2 <sub>1</sub>	P6 <sub>1</sub>
Unit cell (Å)	92.9, 90.0, 93.9	178.4 178.4 247.3
α, β, γ (°)	90.0, 113.3, 90.0	90.0 90.0 120.0
Wavelength (Å)	0.980	0.980
Resolution (Å)	46.6 -2.70 (2.80-2.70)	44.6-3.20 (3.32-3.20)
Unique Reflections	37504 (3085)	73262 (7338)
Completeness (%)	95.7 (79.5)	99.5 (96.9)
Redundancy	3.4 (2.9)	10.3 (9.8)
CC <sub>1/2</sub> (%)	99.8 (91.8)	99.8 (27.7)
<I/σI>	13.14 (2.27)	12.87 (0.56)
R <sub>merge</sub> (%)	6.22 (40.1)	16.4 (492)
R <sub>wp</sub> (%)	3.96 (28.0)	5.36 (165)
Wilson B-factor (Å <sup>2</sup> )	59.53	125.06
<b>Refinement and Validation</b>		
Resolution (Å)	2.70	3.20
Number of atoms	9582	19802
Protein	9516	19696
Ligand	66	106
Waters	0	0

R <sub>work</sub> /R <sub>free</sub> (%)	19.2/24.1	18.8/23.1
R.m.s. deviations		
Bond lengths (Å)	0.009	0.010
Bond angles (°)	1.16	1.23
MolProbity score	1.96	2.04
Clashscore (all atom)	7.61	8.67
Poor rotamers (%)	1.71	1.77
Ramachandran plot		
Favored (%)	94.60	94.36
Allowed (%)	5.07	5.28
Disallowed (%)	0.33	0.36
Average <i>B</i> -factor (Å <sup>2</sup> )	63.16	124.1

<sup>a</sup> Data collected on 12-2 beamline at the Stanford Synchrotron Radiation Lightsource (SSRL).

<sup>b</sup> Numbers in parentheses correspond to the highest resolution shell.

**Table S2. Buried Surface Area (Related to Figure 4)**

Interface Buried Surface Area (Å <sup>2</sup> )				
Structure	C118 Fab/ SARS RBD	C022 Fab/ SARS2 RBD	CR3022 Fab/ SARS2 RBD	COVA1-16 Fab/ SARS2 RBD
PDB	this study	this study	PDB 6W41	PDB 7JMW
<b>Heavy Chain Paratope</b>				
<b>FWRH1</b>	750	769	593	673
<b>CDRH1</b>	0	12	4	0
<b>CDRH2</b>	197	218	271	109
<b>FWRH2</b>	92	3	113	0
<b>FWRH3</b>	461	537	202	564
<b>FWRH4</b>	0	0	0	0
<b>Light Chain Paratope</b>				
<b>CDRL1</b>	272	200	432	154
<b>CDRL2</b>	0	0	0	0
<b>CDRL3</b>	197	130	28	120
<b>CDRL4</b>	92	20	62	0
<b>Total Paratope</b>	1022	969	1025	827
<b>Heavy Chain Epitope</b>	656	704	593	628
<b>Light Chain Epitope</b>	292	202	398	153
<b>Total Epitope</b>	948	906	990	780

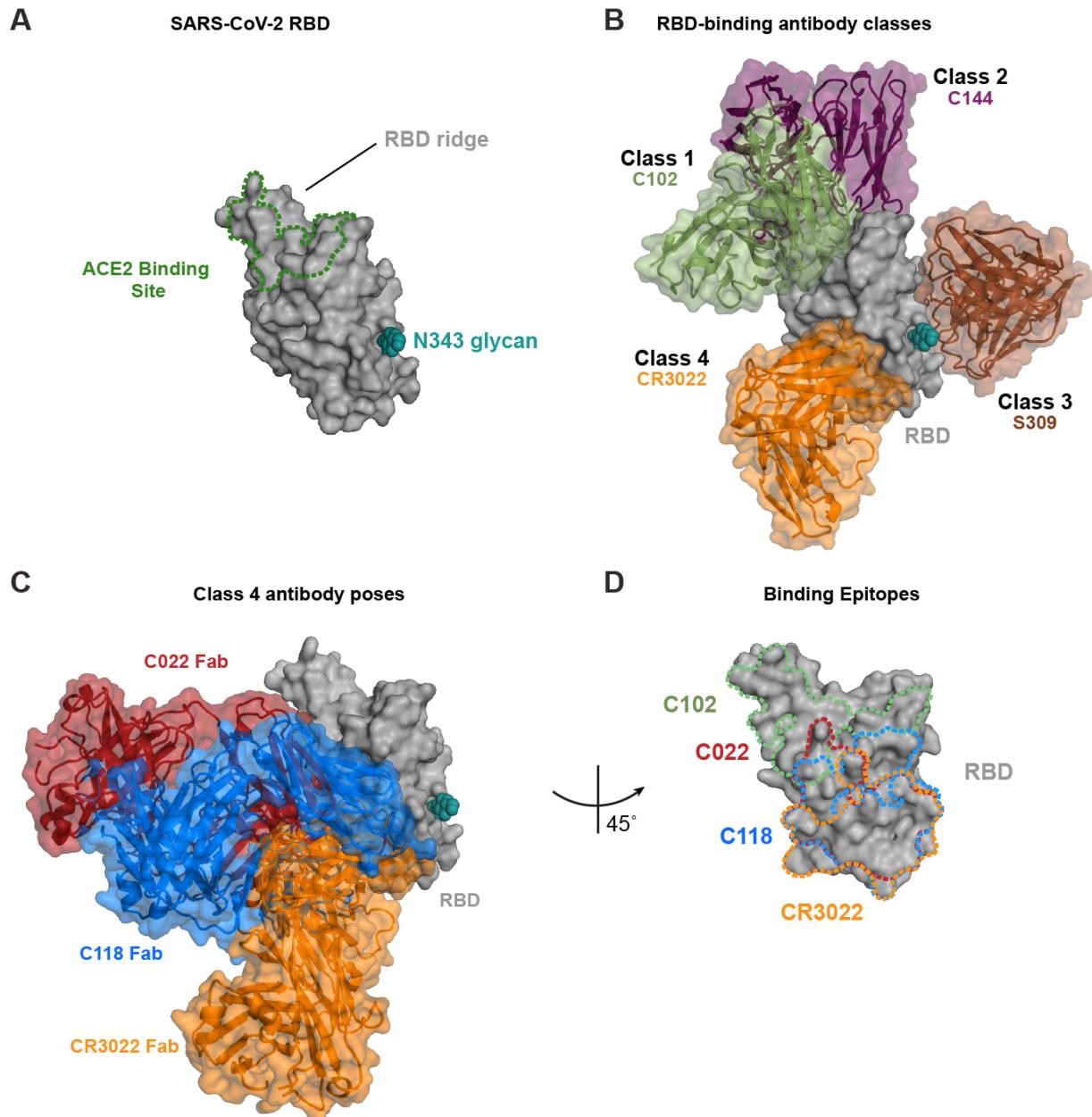


**Table S3. Cryo-EM data collection and refinement statistics for C118-S complex structure (Related to Figure 5).**

	<b>C118</b> <b>SARS-CoV-2 S2P</b> <b>(state 1)</b>	<b>C118</b> <b>SARS-CoV-2 S2P</b> <b>(state 2)</b>
<b>PDB</b>		
<b>EMD</b>		
<b>Data collection conditions</b>		
Microscope	Talos Arctica	
Camera	Gatan K3 Summit	
Magnification	45,000x	
Voltage (kV)	200	
Recording mode	counting	
Dose rate (e <sup>-</sup> /pixel/s)	13.5	
Electron dose (e <sup>-</sup> /Å <sup>2</sup> )	60	
Defocus range (μm)	0.7-2.0	
Pixel size (Å)	0.836	
Micrographs collected	2,970	
Micrographs used	2,487	
Total extracted particles	923,707	
Refined particles	53,728	31,422
Symmetry imposed	C3	C1
Nominal Map Resolution (Å)		
FSC 0.143 (unmasked/masked)	4.4/3.4	7.7/4.5
FSC 0.143 local (unmasked/masked)	4.8/3.7	NA
<b>Refinement and Validation</b>		
Initial model used	7K43	
Number of atoms		
Protein	28,865	
Ligand	795	
MapCC (global/local)	0.83/0.82	
Map sharpening B-factor	88	
R.m.s. deviations		
Bond lengths (Å)	0.01	
Bond angles (°)	0.81	
MolProbity score	2.56	
Clashscore (all atom)	16.5	
Poor rotamers (%)	0	
Ramachandran plot		
Favored (%)	92.6	
Allowed (%)	6.9	
Disallowed (%)	0.5	

250

250



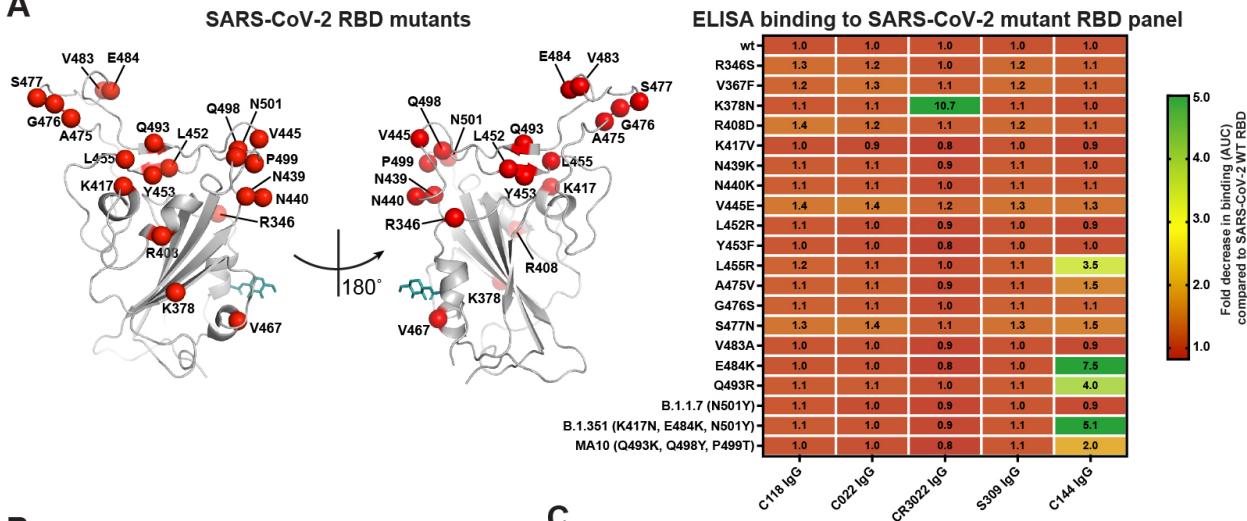
**Figure S1. Epitopes of class 1 – class 4 anti-RBD antibodies (related to Figures 1 and 2)**

(A) SARS-CoV-2 RBD surface representation (grey) with N343 glycan (teal).  
 (B) SARS-CoV-2 RBD surface representation (grey) with overlayed bound models of  $V_HV_L$  of antibodies for class 1 (C102, green, PDB: 7K8M), class 2 (C144, purple, PDB: 7K90), Class 3 (S309, brown, 7JMX), and class 4 (CR3022, orange, PDB: 6W41).

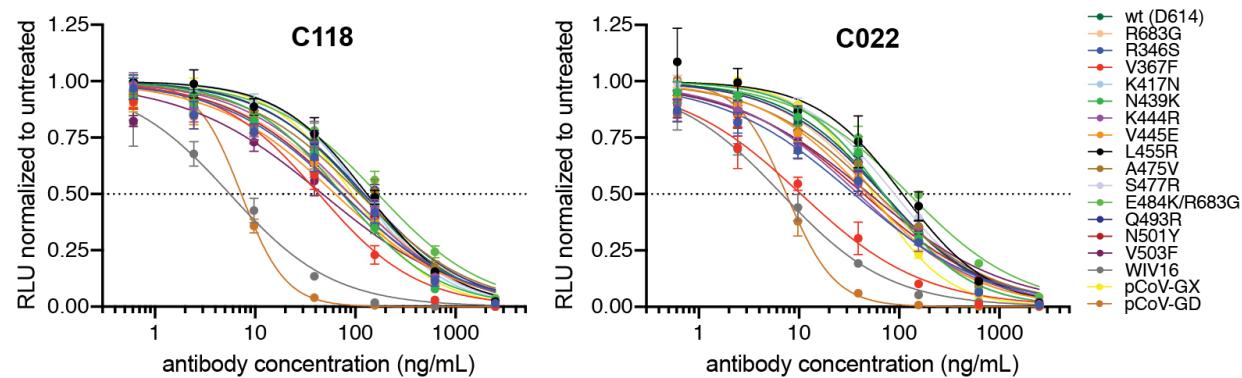
(C) SARS-CoV-2 surface representation with overlay of bound C118 Fab (blue), C022 Fab (red), and CR3022 Fab (orange, PDB: 6W41).

(D) SARS-CoV-2 RBD surface representation (grey) at 45° angle from previous panels. Dotted outlines show epitopes of C102 (green), C022 (red), C118 (blue), and CR3022 (orange) mapped onto SARS-CoV-2 RBD.

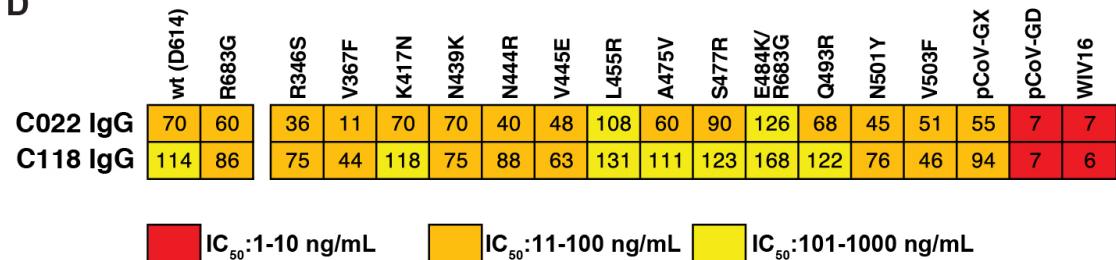
**A**



**B**



**D**

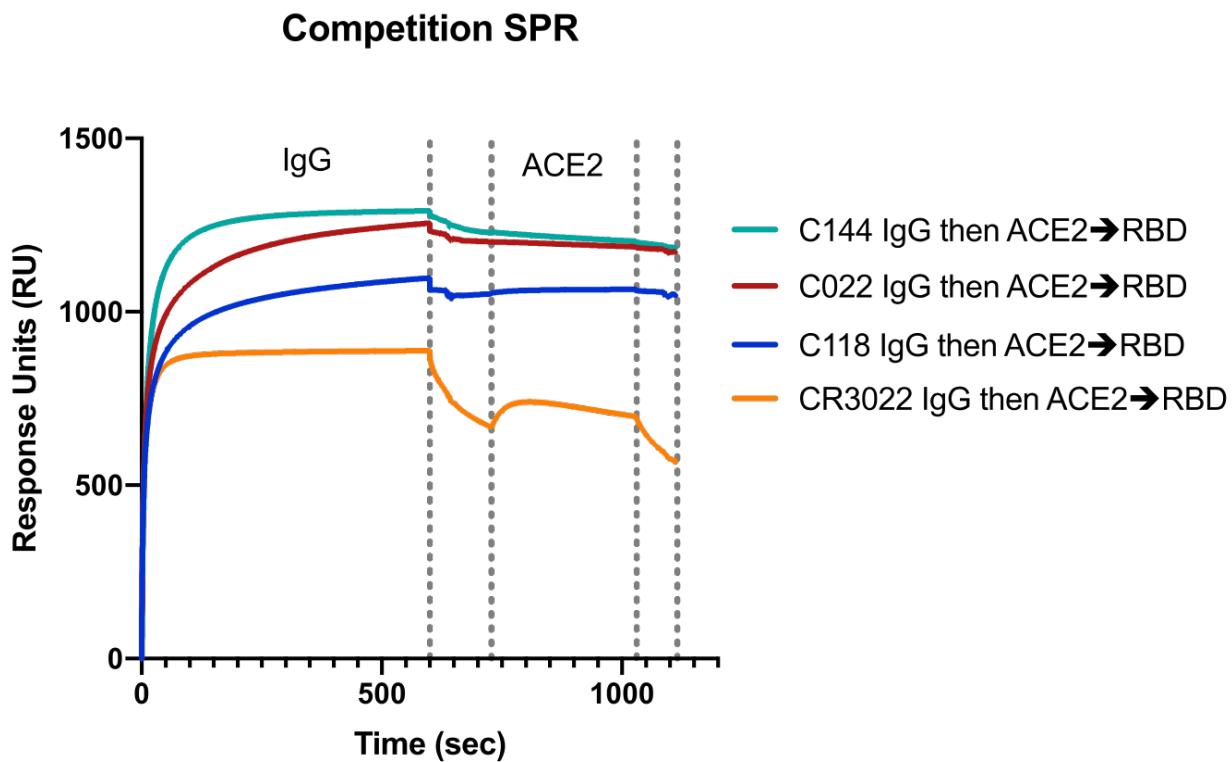


**Figure S2. Binding and Neutralization of SARS-CoV-2 RBD or pseudoviruses carrying spike amino acid substitutions as well as other sarbecovirus pseudoviruses by C118 and C022 (related to Figure 1).**

(A) Left: Cartoon model of SARS-CoV-2 RBD (RBD from C022-RBD structure) showing locations of point mutations as red spheres and the N343<sub>RBD</sub> N-glycan as teal sticks. Right: Comparison of binding of the indicated monoclonal IgGs to RBD mutants from ELISA data shown as AUC values normalized to antibody binding to 'wt' SARS-CoV-2 RBD. Data presented are normalized mean AUC values from two independent experiments.

(B-C) Normalized relative luminescence values for cell lysates of HT1080<sub>ACE2</sub> cells 48h after infection with SARS-CoV-2 pseudovirus carrying indicated spike variants in the presence of increasing concentrations of monoclonal IgGs C118 (B) and C022 (C). The mutants represented substitutions found in circulating SARS-CoV-2 sequences with frequencies >0.01% in GISAID (Shu and McCauley, 2017). Mean and standard deviation of two experiments, each performed in duplicate (n=4), is shown.

(D) Half-maximal inhibitory concentrations (IC<sub>50</sub>) calculated from the neutralization curves in panels B and C for monoclonal IgGs C022 and C118 for neutralization of 'wt' (D614 S trimer) and the indicated mutant SARS-CoV-2 S pseudotyped viruses, as well as other sarbecovirus pseudoviruses. IC<sub>50</sub> values are means of 2 independent experiments. Colors indicate IC<sub>50</sub> ranges, as indicated. The E484K substitution was constructed in an R683G (furin cleavage site mutant) background to increase infectivity (Muecksch et al., 2021).



**Figure S3. SPR-based competition experiment.**

SARS-CoV-2 RBD was coupled to a biosensor chip using primary amine chemistry. An IgG was injected first (seconds 0-600). Seconds 600 - 730 represent the delay required to switch samples for a subsequent injection. From seconds 730 – 1030, soluble ACE2 was injected. Buffer was injected after 1030 seconds.

A

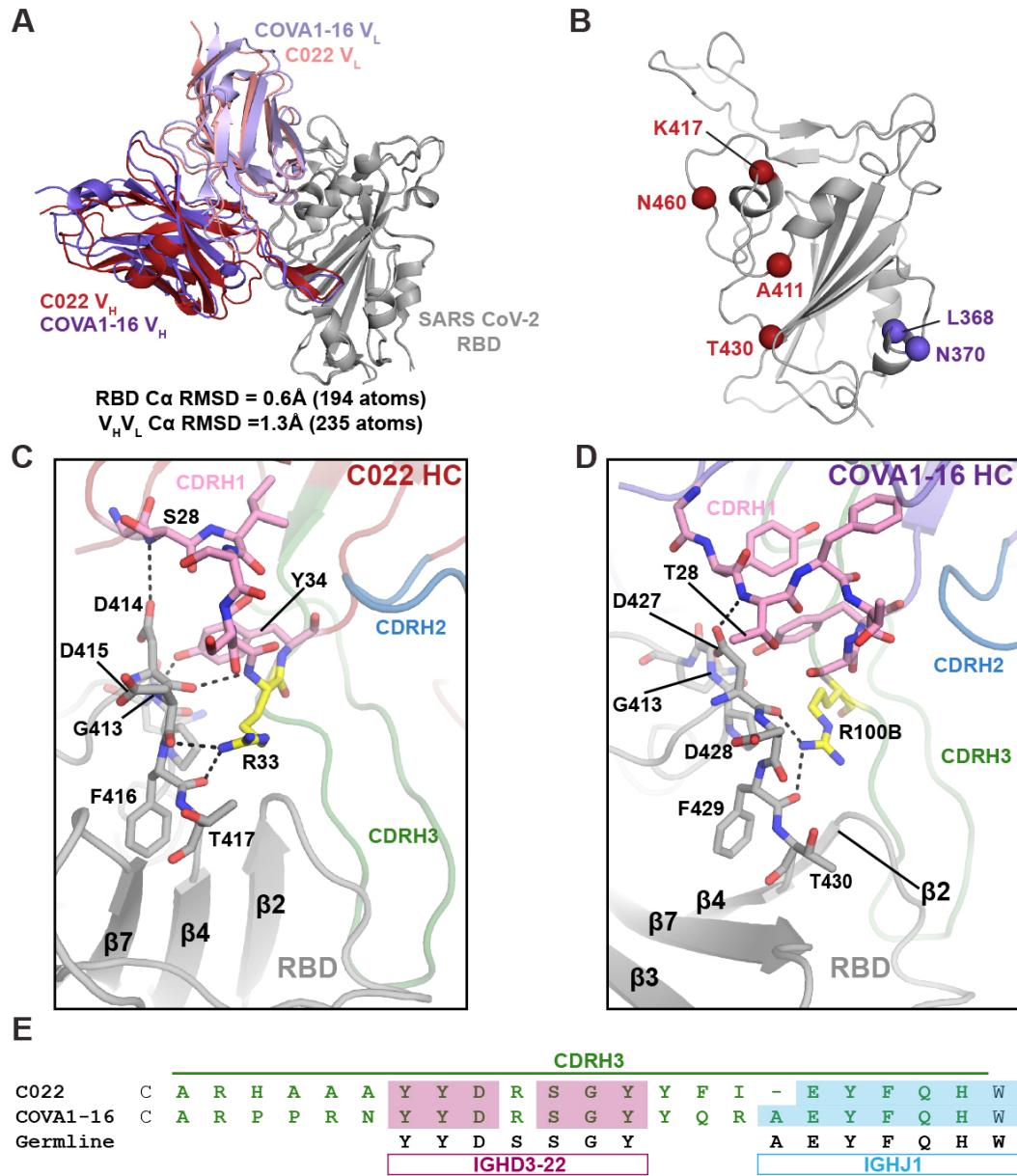
B

	<b>FWR1</b>	<b>CDR1</b>	<b>FWR2</b>	<b>CDR2</b>
C118 V <sub>L</sub>	QPVLTQSPSA-SASLGASVAKLTCTLSSGHSSYIA	WHDQQPEKGPRLMK	LNTDGS	HSKGDD
C022 V <sub>L</sub>	DIQMTQSPSTLSASVGDSVTITCRASQSISWW-	LAWYQQKPGKAPKLIIY	---	KASSLESG
COVA1-16 V <sub>L</sub>	DIQLTQSPSSLSASVGDRVTITCQASQDISNY-	LNWYQQRPGKAPKLIIY	---	DAASNLETG

C

	<i>IGHD3-22</i>	<i>IGHJ1</i>
Germline	ATTACTATGATAGTAGTGTTATTAA	GCTGAATACTTCCAGCACTGG
COVA1-16	TGTGCGAGGCCCTCGAAATTACTATGATAGGAGTGTTATTATCAGAGG	GCTGAATACTTCCAGCACTGG
C022	CTGTGCGAGACATGGCAGCA-TACTATGATAGAAGTGTTATTATTCATC---	GAATACTTCCAGCACTGG

**Figure S4. Comparison of C022, C118, and COVA1-16  $V_H$  $V_L$  sequences (related to Figure 3)**  
 (A,B) Alignment of C118, C022, and COVA1-16 (A)  $V_H$  and (B)  $V_L$  domains. Framework regions (FWRs) and complementarity determining regions (CDRs) assigned using the IMGT definition (Le-franc et al., 2015). Conserved residues (\*), residues with similar properties (:), residues with weakly similar properties (.).  
 (C) Alignment of portions of the C022 and COVA1-16  $V_H$  domain nucleotide sequences to germline gene segment sequences for *IGHD3-22* (sand) and *IGHJ1* (teal).



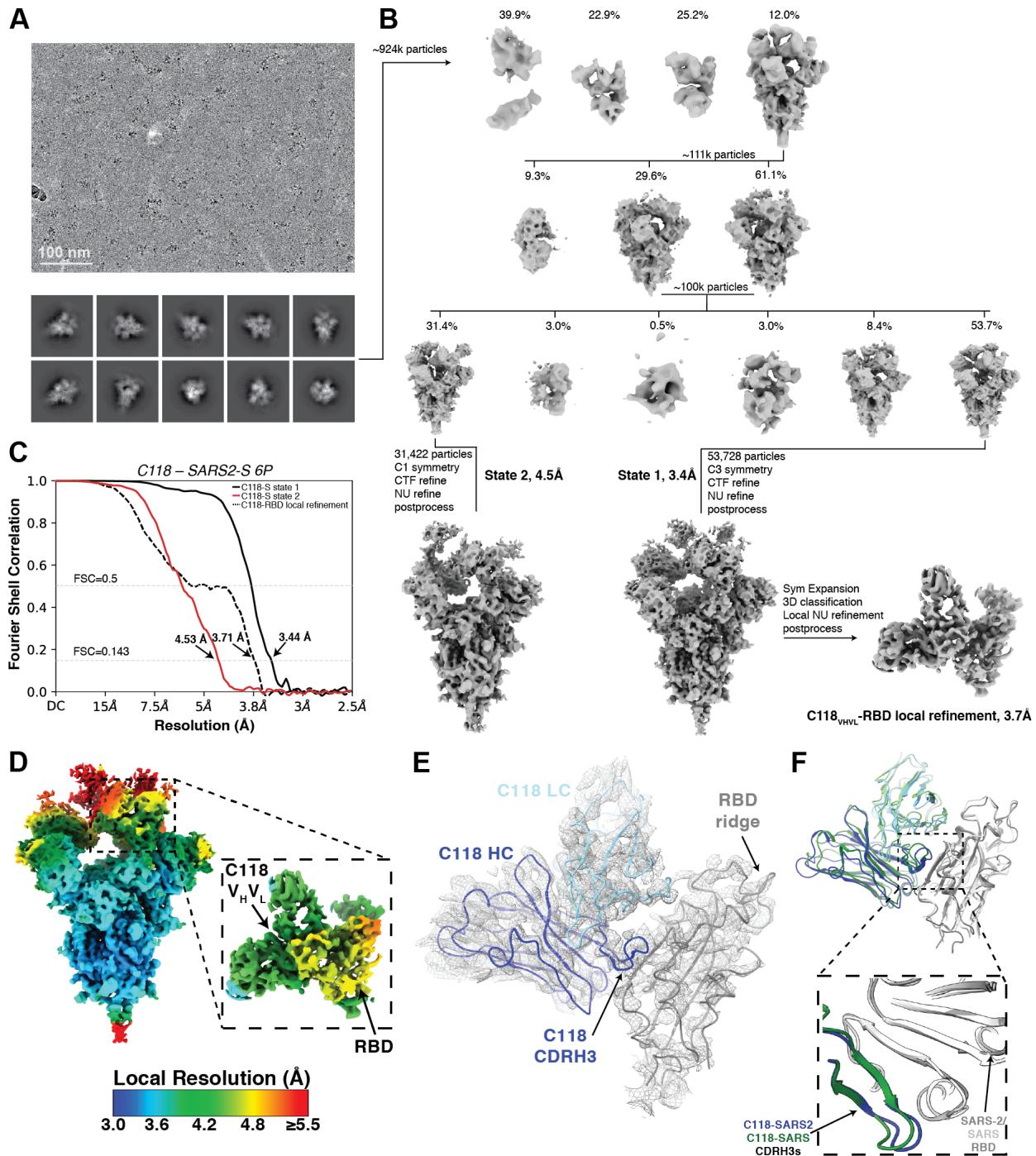
**Figure S5. Comparison of C022 and COVA1-16 antibody binding**

(A) Cartoon representation of superposition of RBDs from C022-RBD and COVA1-16-RBD (PDB 7JMW) crystal structures aligned by their C $\alpha$  atoms.

(B) SARS-CoV-2 RBD cartoon model showing differences between C022 (red) and COVA1-16 (purple) epitopes.

(C,D) Interacting residues of SARS-CoV-2 RBD with the CDRH1 regions from (D) C022 Fab and (E) COVA1-16 Fab (PDB 7JMW). Colors: Oxygens (red), nitrogens (blue). Carbon atoms of critical arginines from each paratope are highlighted in yellow.

(E) Alignment of CDRH3 amino acid sequences for C022 and COVA1-16 and germline-encoded amino acids derived from *IGD3*-22 (mauve) and *IGJ1* (blue) gene segments. Identities between C022 or COVA1-16 with *IGHD3*-22 and *IGHJ1* amino acid sequences shown as shaded boxes and residues within the CDRH3 loop are green.



**Figure S6. Details for cryo-EM C118-S reconstructions (related to Figure 5)**

(A) C118-S 6P representative micrograph (top) and 2D class averages (bottom).

(B) Data processing pipeline for C118-S structure and focused refinement of C118-RBD portion of the structure.

- (C) Gold Standard FSC for final reconstruction of C118-S reconstructions. Resolutions at FSC=0.143 are shown for each volume.
- (D) Local resolution estimates for C118-S map and C118-RBD focused map calculated in cryo-SPARC.
- (E) Ribbon representation of C118-RBD rigid body-refined model into cryoEM density contoured at  $8\sigma$  (gray mesh).
- (F) Overlay of C118-RBD (SARS-CoV-2) from Fab-S cryo-EM structure (blue and gray cartoon) and C118-RBD (SARS-CoV) (green and light gray). CDRH3 interactions are highlighted.

Appendix B: Cryo-EM structure of a CD4-bound open HIV-1 Envelope trimer reveals structural rearrangements of the gp120 V1V2 loop

This appendix describes the Cryo-EM structure of an partially-open CD4-bound HIV Env. Haoqing Wang was the lead author of this project responsible for designing the study preparing the reagents and processing and analyzing the data. My contribution was to do some of the biochemical studies characterizing the interaction between BG505 SOSIP and the 17b antibody that can only bind the open conformation of the Env. I designed that experiment, prepared those reagents, and analyzed the data,

This work was published as:

Wang, H., **Cohen, A.A.**, Galimidi, R.P., Gristick, H.B., Jensen, G.J., and Bjorkman, P.J. (2016). Cryo-EM structure of a CD4-bound open HIV-1 envelope trimer reveals structural rearrangements of the gp120 V1V2 loop. *Proc National Acad Sci* *113*, E7151–E7158.

### **Significance Statement**

The HIV-1 Env glycoprotein exists in multiple conformations on virion surfaces. Although the closed Env state is well characterized, less is known about open Env conformations stabilized by host receptor (CD4) binding. We solved an 8.9Å structure of a partially-open CD4-bound Env trimer by single particle cryo-EM. In the CD4-bound Env, the gp120 V1V2 loops were displaced by ~40Å from their positions at the trimer apex. The displaced V1V2 loops were at the sides of the open trimer in positions adjacent to, and interacting with, the three bound CD4s. These results are relevant to understanding CD4-induced conformational changes leading to co-receptor binding and fusion, HIV-1 Env conformational dynamics, and describe a target structure relevant to drug design and vaccine efforts.

**ABSTRACT**

The HIV-1 envelope (Env) glycoprotein, a trimer of gp120-gp41 heterodimers, relies on conformational flexibility to function in fusing the viral and host membranes. Fusion is achieved after gp120 binds to CD4, the HIV-1 receptor, and a co-receptor, capturing an open conformational state in which the fusion machinery on gp41 gains access to the target cell membrane. In the well-characterized closed Env conformation, the gp120 V1V2 loops interact at the apex of the Env trimer. Less is known about the structure of the open CD4-bound state, in which the V1V2 loops must rearrange and separate to allow access to the coreceptor binding site. We identified two anti-HIV-1 antibodies, the co-receptor mimicking antibody 17b and the gp120-gp41 interface-spanning antibody 8ANC195, that can be added as Fabs to a soluble native-like Env trimer to stabilize it in a CD4-bound conformation. Here we present an 8.9 Å cryo-electron microscopy structure of a BG505 Env-sCD4-17b-8ANC195 complex, which reveals large structural rearrangements in gp120, but small changes in gp41, as compared to closed Env structures. The gp120 protomers are rotated and separated in the CD4-bound structure, and the three V1V2 loops are displaced by ~40 Å from their positions at the trimer apex in closed Env to the sides of the trimer in positions adjacent to, and interacting with, the three bound CD4s. These results are relevant to understanding CD4-induced conformational changes leading to co-receptor binding and fusion, HIV-1 Env conformational dynamics, and describe a target structure relevant to drug design and vaccine efforts.

## INTRODUCTION

The HIV-1 envelope (Env) glycoprotein, a trimer of gp120-gp41 heterodimers, mediates recognition of host receptors and fusion of the viral and target cell membranes (1). Structural flexibility of HIV-1 Env is required for its function in membrane fusion; thus Env exists in multiple conformational states on the surface of virions (2). Fusion involves several steps: the gp120 portion of Env trimer first binds to the host receptor CD4 to capture a conformational state of Env that exposes the binding site for an HIV-1 co-receptor (CCR5 or CXCR4), which in turn leads to gp41-mediated fusion of the viral and host cell membranes. CD4-induced conformational changes within the Env trimer are incompletely understood. The binding of soluble CD4 (sCD4) produces little to no changes in the structures of gp120 cores (gp120 monomers with truncations in the N- and C-termini and V1V2 and V3 loops) (3), but results in rotation of the gp120 protomers within virion-bound Env trimers to create an open conformation distinct from the closed conformation of unliganded virion-bound trimers (4). Single particle electron microscopy (EM) structures of recombinant native-like soluble Env gp140 trimers (SOSIPs) confirmed that they can adopt the same closed and open architectures as virion-bound Env trimers (5-7), thus the SOSIP substitutions (introduction of a disulfide bond linking gp120 to gp41 and an Ile->Pro mutation in gp41 (6)) do not appear to prevent transition to the open state. Despite the plethora of recent crystal and EM structures at atomic and near-atomic resolutions of closed Env trimers, most in complex with broadly neutralizing antibodies (bNAbs) (8-19), only low resolution structures derived from cryo-electron tomography of HIV-1 virions have been available for sCD4-bound open Env trimers (4-7).

The closed conformation of HIV-1 Env is stabilized by interactions at the trimer apex mediated by the gp120 V1V2 loop (8-19). In the closed state, the V1V2 region shields the binding site for the co-receptor on the V3 loop (11, 16), but V1V2 interactions with V3 cannot be maintained when the gp120 protomers rotate and separate to create the CD4-bound open conformation. The details of V1V2 rearrangements in the open structure of HIV-1 Env trimer have not been addressed: the V1V2 loops were not localized in the Env trimer used in a cryo-EM single particle structure of an open KNH1144 SOSIP bound to the co-receptor-mimicking antibody 17b (7) or in lower resolution cryo-electron tomography structures of CD4-bound open Env trimers on virions (4, 7). However, computational modeling suggested displacement of V1V2 towards CD4 in CD4-bound Env structures (20, 21), consistent with earlier studies demonstrating involvement of V1V2 in the induction of the epitopes of co-receptor mimic/CD4-induced antibodies such as 17b (22).

A structural description of conformational changes resulting from CD4 binding requires identification of a stable and conformationally homogeneous CD4-Env trimer complex. We previously reported a 16.8 Å negative stain single particle EM reconstruction of the BG505 SOSIP.664 Env trimer bound to sCD4, 17b, and 8ANC195 (18). The two Fab ligands in this complex bind to distinct epitopes on Env. 8ANC195 binds to a region of the gp120-gp41 interface flanked by N-linked glycans attached to Asn<sub>234<sub>gp120</sub></sub> and Asn<sub>276<sub>gp120</sub></sub> (18, 23). 17b, an HIV-1 co-receptor mimic (24), binds to a CD4-induced (CD4i) epitope (25) that comprises part of the gp120 bridging sheet and overlaps with the co-receptor binding site on gp120 (24). Accordingly, this antibody does not bind or neutralize most HIV-1 strains; the exceptions being sensitive Tier 1 primary isolates in which the co-receptor binding site is exposed in the absence of CD4 (26, 27). Because the co-receptor and 17b binding sites are inaccessible in the closed trimer conformation (11, 16), 17b does not bind to the BG505 SOSIP

trimer unless sCD4 is present (6). Since three 17b Fabs cannot be accommodated on a closed Env trimer due to steric clashes (18), a CD4-bound trimer complex might be stabilized by adding 17b and 8ANC195 (to prevent Env closure and rigidify the gp120-gp41 interface, respectively).

Here we report an 8.9 Å structure of the BG505-sCD4-17b-8ANC195 complex (hereafter referred to as Env-sCD4-17b-8ANC) derived by single particle cryo-EM. The higher resolution structure reveals an Env structure with a large CD4-induced rotation of the gp120 protomers and an ~40 Å movement of the V1V2 loop from the apex of the trimer to a position adjacent to sCD4, but relatively minor changes within gp41 from closed trimer structures. Biochemical studies using a BG505 with a truncated V1V2 loop support the model in which sCD4 binding induces V1V2 displacement to expose the co-receptor binding site. These results are relevant to understanding the CD4-induced conformational changes leading to co-receptor binding and fusion that allows HIV-1 entry into <sup>CD4+</sup> target cells, HIV-1 Env conformational dynamics, and present a target structure relevant to drug design and vaccine efforts.

## Results

### Cryo-EM structure determination

BG505 SOSIP.664, a cleaved, soluble native-like HIV-1 Env trimer (6), sCD4 D1D2 (domains 1 and 2 of the CD4 ectodomain), and Fabs from the anti-HIV-1 antibodies 17b and the G52K5 variant of 8ANC195 (23) (hereafter referred to as 8ANC195) were expressed and purified as previously described (18). The Env-sCD4-17b-8ANC complex was prepared for cryoEM by first isolating an Env-sCD4-17b ternary complex by size exclusion chromatography (SEC), adding 8ANC195 Fab, using

SEC again to isolate the quaternary Env-sCD4-17b-8ANC complex, and then plunge freezing the complex onto EM grids (Fig. S1A).

Two independent single particle reconstructions were obtained from 5,175 out of 13,268 total particles and 9,606 out of 19,355 total particles, respectively, at resolutions of  $\sim 8.9$  Å and  $\sim 9.8$  Å (calculated using the 0.143 gold-standard Fourier shell coefficient cutoff criterion) (28) (Fig. 1, 2; Fig. S2, S3). These relatively high resolutions for a sCD4-bound Env structure were confirmed by feature-based criteria including clear definition of gp41 HR1  $\alpha$ -helices (Fig. 1 C) and densities corresponding to BG505 *N*-linked glycans (Fig. 1 D). Regions of the structure that were disordered and/or calculated to be at a lower resolution (29) were areas most distant from the trimer axis of symmetry such as the constant domains ( $C_H$  and  $C_L$ ) of the Fabs, the D2 domain of sCD4, and density identified as the displaced gp120 V1V2 loop (Fig. S2C).

Coordinates from crystal structures of individual components of the Env-sCD4-17b-8ANC complex were fit by rigid body docking into cryo-EM density maps. The coordinates of 8ANC195 (PDB 4P9M) (23), 17b and sCD4 (PDB 1RZJ) (30) were first docked into their corresponding densities, after which the gp41 coordinates from a BG505 trimer structure (PDB 5CEZ) (9) were fit into density. For fitting gp120 densities, we deleted the V1V2 and V3 coordinates from a closed BG505 trimer structure (PDB 5T3X) (10) and then fit the truncated gp120s individually into protomer densities. After fitting the gp120, 17b, and sCD4 coordinates independently, the complex was compared with the crystal structure of a gp120-sCD4-17b complex (PDB 1RZJ) (30), resulting in root mean square deviations (RMSDs) of 1.7 Å for 98  $C\alpha$  atoms in sCD4 D1 and 1.5 Å for 234  $C\alpha$  atoms in the 17b  $V_H$ - $V_L$  domain after superimposing the gp120s. The relatively low RMSDs for the independently fit

sCD4 and 17b  $V_H$ - $V_L$  coordinates demonstrated that the cryoEM reconstruction reproduced known interactions of CD4 and 17b with gp120. The placement of the 8ANC195  $V_H$ - $V_L$  domains within its epitope at the gp120-gp41 interface was also not greatly shifted from its placement in an 8ANC195-BG505 (closed trimer) crystal structure (PDB 5CJX) (18) (RMSD = 2.3 Å for 238 8ANC195  $V_H$ - $V_L$   $C\alpha$  atoms after superimposing the gp120s) (Fig. 3 A).

The fitted coordinates and density maps for the 8.9 Å and 9.8 Å reconstructions showed no major differences except for the positions of the Fab  $C_H$ - $C_L$  domains (which are not rigid with respect to the antigen-binding  $V_H$ - $V_L$  domains) (Fig. S3). Thus analyses were done using the 8.9 Å reconstruction, with comparisons to verify features of interest with the independently-determined 9.8 Å reconstruction (this study) and the previously-described 16.8 Å negative stain reconstruction (18).

### **Comparison of Env trimer conformational states**

The 8.9 Å Env-sCD4-17b-8ANC structure revealed densities for three sCD4, three 17b, and three 8ANC195 Fabs interacting with a three-fold symmetric BG505 Env trimer (Fig. 1 A, B). The BG505 Env in this complex adopts a conformation that is more open than the closed conformation in crystal and EM structures of Env trimers (8-19), but less open than the conformation in low resolution sCD4-bound Env structures (4-7) (Fig. 2A, B; Fig. 3 A) and an ~9 Å cryo-EM reconstruction of the KN1144 SOSIP.681 soluble trimer bound to 17b in the absence of sCD4 (7). The higher resolution and/or improved order of the present reconstruction revealed features that were unresolved in the other open Env structures, including density for several BG505 *N*-linked glycans (e.g., a well-ordered glycan attached to N386<sub>gp120</sub>) and density for gp41 and gp120  $\alpha$ -helices (Fig. 1 C, D; Fig. S4). The localization of gp41 helices allows comparison of the degree of CD4-induced movement of

gp120 versus gp41 in open and closed Env structures. Superposition of the gp120s from the Env-sCD4-17b-8ANC structure, a closed trimeric Env (PDB 5T3X) (10), and a sCD4-bound Bal open Env structure (PDB 3DNO) (4) revealed major differences in gp120 orientations (Fig. 2 *A, B*; Fig. 3 *A*). Fig. 2 *B* shows a progression of gp120 displacement from the relatively closely-spaced gp120s held together by the V1V2 region in the closed trimeric state (left), to the partially-open conformation in the Env-sCD4-17b-8ANC structure with newly-identified V1V2 loop displacements (middle), to the fully open conformation in the Bal-17b structure (in which the V1V2 loops were not localized) (right). By contrast to the large differences in gp120s in the closed, partially-open, and open Env conformations, the gp41 HR1  $\alpha$ -helices positions were relatively unaffected by CD4-induced Env opening (Fig. 2 *A, B, C, D*).

### CD4-induced V1V2 loop displacement

The BG505 portion of the Env-sCD4-17b-8ANC reconstruction shows no density for the V1V2 and V3 loops in their original positions with respect to gp120, indicating structural rearrangements in addition to the rotation of gp120 monomers that are induced by sCD4 binding. We identified a prominent density near sCD4 as the likely location of the rearranged V1V2 based on several criteria: (*i*) the density projects towards sCD4 starting from the center of the gp120  $\beta$ 2 and  $\beta$ 3  $\beta$ -strands from which the V1V2 loop emanates, (*ii*) analogous density is present in independent CD4-bound Env structures (Fig. 3 *B*): the 8.9  $\text{\AA}$  and 9.8  $\text{\AA}$  cryo-EM reconstructions of the Env-sCD4-17b-8ANC complex described here, the 16.8  $\text{\AA}$  negative stain reconstruction of the Env-sCD4-17b-8ANC complex (18), and the  $\sim$ 20  $\text{\AA}$  sCD4-bound Env trimer structures derived from sub-tomogram averaging of virion-bound Env spikes (4, 5), and (*iii*) the density in the 8.9  $\text{\AA}$  reconstruction contacts sCD4 D1 (Fig. 3 *B*, Fig. 5 *B*), consistent with crystal structures of monomeric gp120s showing V1V2 stem

interactions with sCD4 (30, 31). We note that analogous density is not present in the open structure of the KNH1144 SOSIP.681 trimer bound to 17b Fab in the absence of sCD4 (7), suggesting that the rearranged V1V2 loop becomes more ordered in open Env structures through interactions with bound sCD4.

Resolution limitations in the Env-sCD4-17b-8ANC complex structure precluded ab initio building of V1V2 residues into EM density. However, we could use available V1V2 coordinates to interpret the density since evidence suggests that the V1V2 loop is likely to maintain its overall four-stranded Greek key  $\beta$ -sheet folding topology because this fold is preserved in closed Env trimer structures (8-19) and in structures of V1V2-alone scaffolds (32, 33) (Fig. 4 A). In addition, EM reconstructions of V1V2-directed bNAbs bound to full-length gp120 monomers showed a variety of binding orientations for V1V2 conformation-specific bNAbs (32), consistent with at least some elements of the disulfide-bonded V1V2  $\beta$ -sheet structure being maintained despite flexibility between monomeric gp120 and V1V2. In closed Env trimer structures, strand A of the four-stranded Greek key  $\beta$ -sheet structure of the V1V2 loop emanates a region that includes a helical turn (Fig. 4 B), but the analogous residues are within a  $\beta$ -strand called  $\beta$ 2 in gp120 cores (Fig. 4 C). Strand D, the final  $\beta$ -strand of the V1V2 loop, leads into the gp120  $\beta$ 3  $\beta$ -strand. The environment of the  $\beta$ 2- $\beta$ 3 region in the closed Env trimer is rearranged in sCD4-bound gp120 core structures into a four-stranded antiparallel  $\beta$ -sheet comprising strands  $\beta$ 20,  $\beta$ 21,  $\beta$ 2 and  $\beta$ 3 (Fig. 4 B, C). A molecular dynamics model of repositioned V1V2 in full-length CD4-bound gp120 assumed this same rearrangement of the  $\beta$ 2- $\beta$ 3 region (21) (Fig. 4 A, D).

We used a lower contour level for interpreting the V1V2 density (light red densities in Fig. 2, 3) in our EM maps than we used for central portions of the BG505 trimer; the lower level was required to reveal density for less ordered portions of the complex structure such as sCD4 D2 (Fig. S2C). At a high contour level, we saw that coordinates for the gp120 and sCD4 D1 from a crystal structure of a V1V2-truncated monomeric gp120 core complexed with sCD4 and 17b (PDB 1RZJ) fit the density well (Fig. 5 A). The electron density suggested contacts between V1V2 and sCD4 D1 (Fig. 5 B), consistent with monomeric gp120-sCD4 crystal structures (30, 31). At a lower contour level, we found that the coordinates for the molecular dynamics model of full-length gp120 with a rearranged V1V2 loop fit the EM density well (Fig. 5 C, D). We were unable to localize the V3 loop in the EM density of the partially-open sCD4-bound Env trimer, but we could rule out the location predicted in the molecular dynamics model or in a crystal structure of a V3 loop-containing monomeric gp120 core (PDB 2QAD) (34) (Fig. 5 C; Fig. S5A), perhaps because the V3 loop position in the crystal structure was influenced by crystal packing (Fig. S5B).

### **The role of V1V2 in 17b binding**

To further investigate the influence of the gp120 V1V2 loop on interactions with sCD4, we constructed a V1V2-truncated version of BG505 SOSIP.664 (BG505- $\Delta$ V1V2) analogous to a V1V2-truncated gp120 core (22, 31) with the goal of solving the structure of a V1V2-truncated Env trimer bound to sCD4. Purified BG505- $\Delta$ V1V2 appeared trimeric by negative stain EM, and the complex of BG505- $\Delta$ V1V2 with sCD4, 17b, and 8ANC195 was stable by SEC (Fig. S6A). Trimeric BG505- $\Delta$ V1V2 and some of the individual ligands could be identified in negative stain EM 2D class averages (Fig. S6C). However, we could not derive a 3D reconstruction from the class averages to examine structural differences in BG505 resulting from V1V2 truncation, suggesting that the

BG505-ΔV1V2 complex with sCD4, 17b, and 8ANC195, which lacked the sCD4–V1V2 interaction, was less structurally homogeneous than the Env-sCD4-17b-8ANC complex.

We used the BG505-ΔV1V2 protein to investigate the effects of the V1V2 loop on 17b binding. Purified BG505 or BG505-ΔV1V2 proteins were incubated with 17b Fab in the absence of sCD4 and subjected to SEC. The unmodified BG505 showed no complex formation with 17b, consistent with previous studies (6, 35), whereas BG505-ΔV1V2 formed an SEC-stable complex with 17b in the absence of sCD4 (Fig. 6). These results are consistent with structural changes in the V1V2 loop upon sCD4 binding allowing binding of 17b in the V1V2 loop-containing Env trimer.

## DISCUSSION

The HIV-1 Env spike is a conformationally dynamic molecule, both in its native, virion-bound trimeric state, and in soluble native-like SOSIP trimers developed as immunogen candidates that are being used for biochemical and structural studies (36). At least five different conformational states have been identified by EM and/or X-ray crystallography (listed from closed to increasing open conformation categories): (i) closed unliganded and bNAb-bound conformations observed on virions and in SOSIPs (4-19), (ii) unliganded partially-open native-like states observed by negative stain EM for SOSIPs other than BG505 SOSIP.664 (37-39), (iii) partially-open virion-bound Env trimers complexed with the anti-HIV-1 antibodies b12 or A12 (4), (iv) the partially-open sCD4-17b-8ANC195-BG505 structure reported here and in (18), and (v) an open Env conformation induced on virions by sCD4 and 17b binding (4), on BG505 and other SOSIPs by sCD4 and 17b binding (5), or on the KHN1144 SOSIP by binding of either the 17b (7) or Z13e1(40) antibodies alone. At least some of these conformational states identified through static structural studies are likely to exist on virions,

as evidenced by single molecule fluorescence resonance energy transfer (FRET) studies of Env trimers on HIV-1 virions, which revealed unliganded native Env to be intrinsically dynamic (2). Transitions between low-, intermediate-, and high-FRET states were discovered, with the predominating low-FRET state identified as the closed Env trimer conformation, and the intermediate-FRET state (populated almost exclusively from the high-FRET state) interpreted as a co-receptor-stabilized conformation that was stabilized by simultaneous introduction of sCD4 and 17b (2). Although the high-FRET state could not be precisely identified, the proportions of both the high- and medium-FRET states were increased by sCD4 and 17b addition, suggesting they represent distinct forms of open Env conformational states.

The large conformational differences in the sCD4-17b-8ANC195– bound BG505 trimer with respect to closed Envs, including rotation/separation of the gp120 cores and a >40 Å displacement of the V1V2 loop, suggests that this conformation represents a structural intermediate to co-receptor binding that is closer to completely open sCD4-bound Env structures than to closed structures. The sCD4-induced Env trimer opening observed in this study arose primarily from rigid body rotations of the gp120 monomers as opposed to changes in trimeric gp41 or the gp120-gp41 interface. With respect to trimeric gp41, we observed that the HR1  $\alpha$ -helices in the partially-open trimer were slightly closer together than their counterparts in a closed trimer (Fig. 2 B), but cannot rule out artifacts from the low resolution of the partially-open structure and/or effects of the SOSIP ‘IP’ substitution (I559P<sub>gp41</sub>) (6) on sCD4-induced conformation changes. However, the conclusion that the gp120-gp41 interface is relatively unchanged during sCD4-induced trimer opening is supported by the fact that 8ANC195 binds similarly to both closed and sCD4-bound trimers (Fig. 3 A). The question remains as to why the BG505 in our structure is only partially-open as opposed to completely

open. Because the Env-sCD4-17b-8ANC complex was prepared by adding 8ANC195 to a pre-formed BG505-sCD4-17b complex, we speculate that the BG505-sCD4-17b complex was fully open until incubated with 8ANC195, a bNAb that prefers the closed Env conformation (18). 8ANC195 binding could have induced partial closure of the Env trimer to better bind to its gp120-gp41 epitope, a conformational sequence reminiscent of the high- to intermediate-FRET state conversion described for native Env trimers on virions (2). It is notable that the Env trimer did not completely close, which would have resulted in steric occlusion of the three 17b Fabs at the trimer apex (18) likely leading to 17b Fab dissociation, nor did it even close to the point of creating contacts between 17b Fabs (as evidenced by the cryo-EM map showing no contact between 17b Fabs) (Fig. 1 A). Thus the partially-open Env conformation revealed in the Env-sCD4-17b-8ANC likely represents a conformation accessible to native Env trimers on the pathway towards fusion.

In summary, the cryo-EM structure presented here represents the most detailed glimpse of structural changes occurring during Env-mediated fusion of the HIV-1 and host cell membranes. The presence of 17b, a co-receptor mimicking antibody, suggests that the partially-open Env conformation we described is correlated with the co-receptor-bound Env state. Hence the V1V2 loop movement observed in our complex structure rationalizes why CD4 binding is required for co-receptor binding and subsequent release of the fusion peptide. In addition, the structure provides a new potential target for design of antibody- or small molecule-based anti-HIV-1 therapeutics.

## Methods

**Protein Production and Purification.** 6x-His tagged Fabs of 17b and the 8ANC195<sub>G52K5</sub> variant of 8ANC195 were expressed by transient transfection in HEK293-6E cells (National Research Council of Canada) and purified from cell supernatants using Ni-NTA chromatography and SEC as described previously (18). sCD4 D1D2 (domains 1 and 2; residues 1–186 of mature CD4) was produced in baculovirus-infected Hi5 insect cells and was purified by Ni-NTA chromatography and SEC (41). BG505 SOSIP.664, a native-like soluble clade A gp140 trimer (6), was constructed to include ‘SOS’ substitutions (A501C<sub>gp120</sub>, T605C<sub>gp41</sub>), the ‘IP’ substitution (I559P<sub>gp41</sub>), the *N*-linked glycan sequence at residue 332<sub>gp120</sub> (T332N<sub>gp120</sub>), an enhanced gp120-gp41 cleavage site (REKR to RRRRRR), and a stop codon after residue 664<sub>gp41</sub> (Env numbering according to HX nomenclature). BG505-ΔV1V2 trimer was constructed by replacing residues 128-194 of the V1V2 loop with a Gly-Ala-Gly linker, as described for previous experiments with a V1V2-truncated gp120 (22). BG505 and BG505-ΔV1V2 proteins were expressed in HEK293-6E cells treated with 5 μM kifunensine (Sigma) by transient transfection of plasmids encoding Env trimer and soluble furin at a ratio of 4:1 as previously described (18). BG505 SOSIP proteins were isolated from cell supernatants using a 2G12 immunoaffinity column as described (10). After elution with 3M MgCl<sub>2</sub> followed by immediate buffer exchange into Tris-buffered saline pH 8.0 (TBS), trimers were purified by SEC using a Superdex 200 16/60 column, Mono Q ion exchange chromatography, and a second SEC purification using a Superose 6 10/300 column (columns from GE Healthcare).

**Cryo-EM data collection and processing.** The Env-sCD4-17b-8ANC complex was made by incubating BG505 with excess sCD4 and 17b Fab overnight and then isolated by SEC. After incubating the BG505-sCD4-17b complex with excess 8ANC195 Fab for two hours at 4 °C, the quaternary complex Env-sCD4-17b-8ANC complex was isolated by SEC. Purified Env-sCD4-17b-8ANC complexes were diluted to 60 µg/ml in TBS and vitrified in liquid ethane using a Mark IV Vitrobot. Sample grids were prepared by adding 3 µL of complex to glow discharged 400 Mesh Quantifoil ® R1.2/1.3 copper grids (for the 8.9 Å reconstruction) or to 400 Mesh C-Flat™ R1.2/1.3 grids (for the 9.8 Å reconstruction). Images were recorded on a Titan Krios electron microscope equipped with Gatan K2 Summit direct detector and an energy filter with a slit width of 20 eV (for the 9.8 Å reconstruction only) using SerialEM (42). For the 8.9 Å reconstruction, 10 sec exposures were divided into 25 subframes and the dose rate was 3.8 electrons/pixel/subframe. For the 9.8 Å reconstruction, 20.25 sec exposures were divided into 54 subframes and the dosage rate was 3.7 electrons/pixel/subframe. After binning by 2 and motion correction, each image was 4k x 4k and 1.64 Å per pixel (8.9 Å reconstruction) or 4k x 4k and 1.71 Å/pixel (9.8 Å reconstruction).

Both data sets were motion corrected and dose weighted using Unblur and Summovie (43-45). Motion corrected micrographs without dose weighting were used for CTF estimations. Motion corrected micrographs with dose weighting were used for particle picking, and motion corrected micrographs with dose weighting and restored noise power after filtering were used for all classification and refinement processes.

Particles were picked using the SWARM method of EMAN2.1 (46) and CTF estimations were done using CTFFIND4 (47). For the 8.9 Å reconstruction: A total of 808 movies were collected. After

motion correction and dose weighting, CTF curves were confidently fit to beyond 6 Å in 360 micrographs; the others were discarded. A total of 13,268 particles were picked. Particles were classified in 2D with Relion (48), resulting in 130 2D classes. Of these, 9 classes from 6,675 particles were selected as “good” classes. For 3D classification, the 16.8 Å Env-sCD4-17b-8ANC negative stain structure (EMDB 3086) was used as a reference and the C<sub>H</sub>-C<sub>L</sub> domains of the Fabs were masked out; two 3D classes were then produced. After selecting one 3D class as a “good” class, 5,175 particles remained for 3D refinement. After 3D refinement, post processing, particle polishing, and gold-standard FSC estimations were done using Relion (48) following procedures in the tutorial. Density maps were low pass filtered to 5 Å to remove noise. Local resolution estimations were done using ResMap (29).

The 9.8 Å structure was produced in the same way with minor differences: 642 movies were collected; only 480 micrographs could be CTF fit to beyond 6 Å; a total of 19,355 particles were picked; 11,915 were retained after 2D classification; and 9,606 particles retained through 3D classification.

After motion correcting, dose weighting, and CTF correcting the two data sets individually, the CTF-corrected (flipped) particles were scaled to the same Å/pixel value and combined into a single data set of 32,623 particles. 18,476 were retained after 2D classification; and 17,013 particles were retained through 3D classification. The resolution of a reconstruction calculated from the combined data sets was 9.6 Å (Fig. S3C), lower than the 8.9 Å resolution calculated for a reconstruction from the first data set. Because combining the data sets did not improve the resolution beyond 8.9 Å, we kept the data sets separated for structural analyses.

**Model building.** Coordinates from crystal structures were manually fit into cryo-EM density maps as rigid bodies using UCSF Chimera's Fit in Map function (49), and the complex was further refined using real-space refinement in PHENIX (50). Coordinates used for fitting or comparisons were gp120 from BG505 SOSIP.664 (closed conformation) (PDB 5T3X), gp41 from BG505 SOSIP.664 (closed conformation) (PDB 5CEZ or 5T3X), sCD4 (PDB 2NXY), 17b Fab (PDB 2NXY), 8ANC195 Fab (PDB 4P9M), 16.8 Å negative stain EM structure of Env-sCD4-17b-8ANC (EMDB 3086), BG505-8ANC195 (PDB 5CJX) and X1193.c1 SOSIP.664-PGT122-35O22-VRC01 (PDB 5FYJ) closed conformation complexes, V1V2 scaffold (PDB 5ESV), gp120 with V1V2 model from molecular dynamics (coordinates obtained from Hironori Sato), gp120 core-sCD4-17b complex (PDB 1RZJ), gp120 core with ordered V3 loop (2QAD), open 17b-bound KNH1144 SOSIP.681 (EMDB entry 5462; coordinates for gp120s obtained from Sriram Subramaniam), gp120-sCD4-17b complex (PDB 2NXY), open conformation virion-bound Bal Env-sCD4 (EMD: 5455), and open conformation virion-bound Bal Env-sCD4-17b (EMDB 5020, PDB 3DNO).

The contour levels for EM maps in this study were chosen based on local resolution estimations and fitted coordinates. For rigid body docking, model building, and coordinates visualization, we chose a contour level such that the gp41 HR1  $\alpha$ -helices exactly fit into the density. However, density for the D2 domain of sCD4 could not be visualized at this high contour level. We lowered the contour level such that density for the D2 domain of sCD4 appeared and used this lower contour level to interpret the density corresponding to the rearranged V1V2 loop region. For determining the rotation and translation relating the gp120s in the closed and partially-open conformations, the transformation relating gp120  $\alpha$ -helices at positions 60-63, 99-113, 335-349, and 475-483 was calculated

using TM-align (51). The corresponding screw transformation was calculated according to (52) and visualized using AntibodyDatabase (53).

**Negative stain EM.** The BG505-ΔV1V2 -sCD4-17b-8ANC195 quaternary complex was made as described above for the Env-sCD4-17b-8ANC complex. Purified complexes were diluted to 10 ug/ml in TBS immediately before adding 3  $\mu$ l to a glow discharged ultrathin C film on holey carbon support film, 400 mesh, Cu grids (Ted Pella) and staining with uranyl acetate. Data were collected using a FEI Tecnai T12 transmission electron microscope operating at 120 keV with a Gatan Ultrascan 2k  $\times$  2k CCD camera. Each image was collected using a 0.5 s exposure at  $\sim$ 1  $\mu$ m defocus and 42,000x magnification resulting in 2.5  $\text{\AA}$  per pixel. For the BG505-ΔV1V2-sCD4-17b-8ANC195 complex, a total of 7,251 particles were picked and CTF corrected using EMAN2.1 (46). Reference-free 2D classification was performed using Relion (48).

## ACKNOWLEDGEMENTS

We thank Zhiheng Yu, Chuan Hong, and Rick Huang (Janelia Farm) for assistance with cryo-EM data collection and motion correction, Hironori Sato for coordinates of the molecular dynamics model of full-length gp120, Sriram Subramaniam for gp120 coordinates from the KNH1144 SO-SIP.681-17b complex, Alasdair McDowall and Songye Chen for training in cryo-EM techniques and data processing, Anthony West for performing alignment calculations, Jost Vielmetter and the Caltech Protein Expression Center for transfections and protein expression, and members of the Bjorkman and Jensen laboratories for helpful discussions and critical reading of the manuscript. This research was supported by the National Institutes of Health Grant 2 P50 GM082545-06 (P.J.B.), National Institute Of Allergy and Infectious Diseases of the National Institutes of Health Grant

HIVRAD P01 AI100148 (P.J.B.) (the content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health), and the Bill and Melinda Gates Foundation (Collaboration for AIDS Vaccine Discovery Grant 1040753 [P.J.B.]). We thank the Gordon and Betty Moore and Beckman Foundations for gifts to Caltech to support electron microscopy.

Accession Codes. The atomic coordinates have been deposited in the PDB under the accession code XXXX. The cryo-EM reconstruction has been deposited in the Electron Microscopy Data Bank under the accession code EMD-XXXX.

Author contributions: H.W. and P.J.B. designed research; H.W. and A.A.C. performed research; H.W., A.A.C., R.P.G., H.B.G., G.J.J., and P.J.B. analyzed data; and H.W., G.J.J., and P.J.B. wrote the paper with contributions from all authors.

## REFERENCES

1. West AP, Jr., Scharf L, Scheid JF, Klein F, Bjorkman PJ, & Nussenzweig MC (2014) Structural Insights on the Role of Antibodies in HIV-1 Vaccine and Therapy. *Cell* 156(4):633-648.
2. Munro JB, Gorman J, Ma X, Zhou Z, Arthos J, Burton DR, Koff WC, Courter JR, Smith AB, 3rd, Kwong PD, Blanchard SC, & Mothes W (2014) Conformational dynamics of single HIV-1 envelope trimers on the surface of native virions. *Science* 346(6210):759-763.
3. Kwon YD, Finzi A, Wu X, Dogo-Isonagie C, Lee LK, Moore LR, Schmidt SD, Stuckey J, Yang Y, Zhou T, Zhu J, Vicić DA, Debnath AK, Shapiro L, Bewley CA, Mascola JR, Sodroski JG, & Kwong PD (2012) Unliganded HIV-1 gp120 core structures assume the CD4-bound conformation with regulation by quaternary interactions and variable loops. *Proc Natl Acad Sci U S A* 109(15):5663-5668.
4. Liu J, Bartesaghi A, Borgnia MJ, Sapiro G, & Subramaniam S (2008) Molecular architecture of native HIV-1 gp120 trimers. *Nature* 455(7209):109-113.
5. Harris A, Borgnia MJ, Shi D, Bartesaghi A, He H, Pejchal R, Kang YK, Depetris R, Marozsan AJ, Sanders RW, Klasse PJ, Milne JL, Wilson IA, Olson WC, Moore JP, & Subramaniam S (2011) Trimeric HIV-1 glycoprotein gp140 immunogens and native HIV-1 envelope glycoproteins display the same closed and open quaternary molecular architectures. *Proc Natl Acad Sci U S A* 108(28):11440-11445.
6. Sanders RW, Derking R, Cupo A, Julien JP, Yasmeen A, de Val N, Kim HJ, Blattner C, de la Pena AT, Korzun J, Golabek M, de Los Reyes K, Ketas TJ, van Gils MJ, King CR, Wilson IA, Ward AB, Klasse PJ, & Moore JP (2013) A next-generation cleaved, soluble HIV-1 Env Trimer, BG505 SOSIP.664 gp140, expresses multiple epitopes for broadly neutralizing but not non-neutralizing antibodies. *PLoS Pathog* 9(9):e1003618.
7. Tran EE, Borgnia MJ, Kuybeda O, Schauder DM, Bartesaghi A, Frank GA, Sapiro G, Milne JL, & Subramaniam S (2012) Structural mechanism of trimeric HIV-1 envelope glycoprotein activation. *PLoS Pathog* 8(7):e1002797.
8. Bartesaghi A, Merk A, Borgnia MJ, Milne JL, & Subramaniam S (2013) Prefusion structure of trimeric HIV-1 envelope glycoprotein determined by cryo-electron microscopy. *Nat Struct Mol Biol* 20(12):1352-1357.
9. Garces F, Lee JH, de Val N, Torrents de la Pena A, Kong L, Puchades C, Hua Y, Stanfield RL, Burton DR, Moore JP, Sanders RW, Ward AB, & Wilson IA (2015) Affinity Maturation of a Potent Family of HIV Antibodies Is Primarily Focused on Accommodating or Avoiding Glycans. *Immunity* 43(6):1053-1063.
10. Gristick HB, von Boehmer L, West AP, Jr., Schamber M, Gazumyan A, Golijanin J, Seaman MS, Fatkenheuer G, Klein F, Nussenzweig MC, & Bjorkman PJ (2016) Natively glycosylated HIV-1 Env structure reveals new mode for antibody recognition of the CD4-binding site. *Nat Struct Mol Biol*:in press.
11. Julien JP, Cupo A, Sok D, Stanfield RL, Lyumkis D, Deller MC, Klasse PJ, Burton DR, Sanders RW, Moore JP, Ward AB, & Wilson IA (2013) Crystal structure of a soluble cleaved HIV-1 envelope trimer. *Science* 342(6165):1477-1483.

12. Kong L, Torrents de la Pena A, Deller MC, Garces F, Sliepen K, Hua Y, Stanfield RL, Sanders RW, & Wilson IA (2015) Complete epitopes for vaccine design derived from a crystal structure of the broadly neutralizing antibodies PGT128 and 8ANC195 in complex with an HIV-1 Env trimer. *Acta Crystallogr D Biol Crystallogr* 71(Pt 10):2099-2108.
13. Kwon YD, Pancera M, Acharya P, Georgiev IS, Crooks ET, Gorman J, Joyce MG, Guttman M, Ma X, Narpala S, Soto C, Terry DS, Yang Y, Zhou T, Ahlsen G, Bailer RT, Chambers M, Chuang GY, Doria-Rose NA, Druz A, Hallen MA, Harned A, Kirys T, Louder MK, O'Dell S, Ofek G, Osawa K, Prabhakaran M, Sastry M, Stewart-Jones GB, Stuckey J, Thomas PV, Tittley T, Williams C, Zhang B, Zhao H, Zhou Z, Donald BR, Lee LK, Zolla-Pazner S, Baxa U, Schon A, Freire E, Shapiro L, Lee KK, Arthos J, Munro JB, Blanchard SC, Mothes W, Binley JM, McDermott AB, Mascola JR, & Kwong PD (2015) Crystal structure, conformational fixation and entry-related interactions of mature ligand-free HIV-1 Env. *Nat Struct Mol Biol* 22(7):522-531.
14. Lee JH, de Val N, Lyumkis D, & Ward AB (2015) Model Building and Refinement of a Natively Glycosylated HIV-1 Env Protein by High-Resolution Cryo-electron Microscopy. *Structure* 23(10):1943-1951.
15. Lee JH, Ozorowski G, & Ward AB (2016) Cryo-EM structure of a native, fully glycosylated, cleaved HIV-1 envelope trimer. *Science* 351(6277):1043-1048.
16. Lyumkis D, Julien JP, de Val N, Cupo A, Potter CS, Klasse PJ, Burton DR, Sanders RW, Moore JP, Carragher B, Wilson IA, & Ward AB (2013) Cryo-EM structure of a fully glycosylated soluble cleaved HIV-1 envelope trimer. *Science* 342(6165):1484-1490.
17. Pancera M, Zhou T, Druz A, Georgiev IS, Soto C, Gorman J, Huang J, Acharya P, Chuang GY, Ofek G, Stewart-Jones GB, Stuckey J, Bailer RT, Joyce MG, Louder MK, Tumba N, Yang Y, Zhang B, Cohen MS, Haynes BF, Mascola JR, Morris L, Munro JB, Blanchard SC, Mothes W, Connors M, & Kwong PD (2014) Structure and immune recognition of trimeric pre-fusion HIV-1 Env. *Nature* 514(7523):455-461.
18. Scharf L, Wang H, Gao H, Chen S, McDowall AW, & Bjorkman PJ (2015) Broadly Neutralizing Antibody 8ANC195 Recognizes Closed and Open States of HIV-1 Env. *Cell* 162(6):1379-1390.
19. Stewart-Jones GBE, Soto C, Lemmin T, Chuang G-Y, Druz A, Kong R, Thomas PV, Wagh K, Zhou T, Behrens A-J, Bylund T, Choi CW, Davison JR, Georgiev IS, Joyce MG, Do Kwon Y, Pancera M, Taft J, Yang Y, Zhang B, Shivatare SS, Shivatare VS, Lee C-CD, Wu CY, Bewley CA, Burton DR, Koff WC, Connors M, Crispin M, Baxa U, Korber BT, Wong CH, Mascola JR, & Kwong PD (2016) Trimeric HIV-1-Env Structures Define Glycan Shields from Clades A, B, and G. *Cell* 165:813-826.
20. Rasheed M, Bettadapura R, & Bajaj C (2015) Computational Refinement and Validation Protocol for Proteins with Large Variable Regions Applied to Model HIV Env Spike in CD4 and 17b Bound State. *Structure* 23(6):1138-1149.
21. Yokoyama M, Nomaguchi M, Doi N, Kanda T, Adachi A, & Sato H (2016) In silico Analysis of HIV-1 Env-gp120 Reveals Structural Bases for Viral Adaptation in Growth-Restrictive Cells. *Front Microbiol* 7:110.
22. Wyatt R, Moore J, Accola M, Desjardin E, Robinson J, & Sodroski J (1995) Involvement of the V1/V2 variable loop structure in the exposure of human immunodeficiency virus type 1 gp120 epitopes induced by receptor binding. *J Virol* 69(9):5723-5733.
23. Scharf L, Scheid JF, Lee JH, West AP, Jr., Chen C, Gao H, Gnanapragasam PN, Mares R, Seaman MS, Ward AB, Nussenzweig MC, & Bjorkman PJ (2014) Antibody 8ANC195

reveals a site of broad vulnerability on the HIV-1 envelope spike. *Cell reports* 7(3):785-795.

24. Kwong PD, Wyatt R, Robinson J, Sweet RW, Sodroski J, & Hendrickson WA (1998) Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature* 393(6686):648-659.
25. Thali M, Moore JP, Furman C, Charles M, Ho DD, Robinson J, & Sodroski J (1993) Characterization of conserved human immunodeficiency virus type 1 gp120 neutralization epitopes exposed upon gp120-CD4 binding. *J Virol* 67(7):3978-3988.
26. Salzwedel K, Smith ED, Dey B, & Berger EA (2000) Sequential CD4-coreceptor interactions in human immunodeficiency virus type 1 Env function: Soluble CD4 activates Env for coreceptor-dependent fusion and reveals blocking activities of antibodies against cryptic conserved epitopes on gp120. *Journal of Virology* 74(1):326-333.
27. Seaman MS, Janes H, Hawkins N, Grandpre LE, Devoy C, Giri A, Coffey RT, Harris L, Wood B, Daniels MG, Bhattacharya T, Lapedes A, Polonis VR, McCutchan FE, Gilbert PB, Self SG, Korber BT, Montefiori DC, & Mascola JR (2010) Tiered categorization of a diverse panel of HIV-1 Env pseudoviruses for assessment of neutralizing antibodies. *J Virol* 84(3):1439-1452.
28. Scheres SH & Chen S (2012) Prevention of overfitting in cryo-EM structure determination. *Nat Methods* 9(9):853-854.
29. Kucukelbir A, Sigworth FJ, & Tagare HD (2014) Quantifying the local resolution of cryo-EM density maps. *Nat Methods* 11(1):63-65.
30. Huang CC, Venturi M, Majeed S, Moore MJ, Phogat S, Zhang MY, Dimitrov DS, Hendrickson WA, Robinson J, Sodroski J, Wyatt R, Choe H, Farzan M, & Kwong PD (2004) Structural basis of tyrosine sulfation and VH-gene usage in antibodies that recognize the HIV type 1 coreceptor-binding site on gp120. *Proc Natl Acad Sci U S A* 101(9):2706-2711.
31. Zhou T, Xu L, Dey B, Hessell AJ, Van Ryk D, Xiang SH, Yang X, Zhang MY, Zwick MB, Arthos J, Burton DR, Dimitrov DS, Sodroski J, Wyatt R, Nabel GJ, & Kwong PD (2007) Structural definition of a conserved neutralization epitope on HIV-1 gp120. *Nature* 445(7129):732-737.
32. McLellan JS, Pancera M, Carrico C, Gorman J, Julien JP, Khayat R, Louder R, Pejchal R, Sastry M, Dai K, O'Dell S, Patel N, Shahzad-ul-Hussan S, Yang Y, Zhang B, Zhou T, Zhu J, Boyington JC, Chuang GY, Diwanji D, Georgiev I, Kwon YD, Lee D, Louder MK, Moquin S, Schmidt SD, Yang ZY, Bonsignori M, Crump JA, Kapiga SH, Sam NE, Haynes BF, Burton DR, Koff WC, Walker LM, Phogat S, Wyatt R, Orwenyo J, Wang LX, Arthos J, Bewley CA, Mascola JR, Nabel GJ, Schief WR, Ward AB, Wilson IA, & Kwong PD (2011) Structure of HIV-1 gp120 V1/V2 domain with broadly neutralizing antibody PG9. *Nature* 480(7377):336-343.
33. Gorman J, Soto C, Yang MM, Davenport TM, Guttman M, Bailer RT, Chambers M, Chuang GY, DeKosky BJ, Doria-Rose NA, Druz A, Ernandes MJ, Georgiev IS, Jarosinski MC, Joyce MG, Lemmin TM, Leung S, Louder MK, McDaniel JR, Narpala S, Pancera M, Stuckey J, Wu X, Yang Y, Zhang B, Zhou T, Program NCS, Mullikin JC, Baxa U, Georgiou G, McDermott AB, Bonsignori M, Haynes BF, Moore PL, Morris L, Lee KK, Shapiro L, Mascola JR, & Kwong PD (2016) Structures of HIV-1 Env V1V2 with broadly neutralizing antibodies reveal commonalities that enable vaccine design. *Nat Struct Mol Biol* 23(1):81-90.

34. Huang CC, Tang M, Zhang MY, Majeed S, Montabana E, Stanfield RL, Dimitrov DS, Korber B, Sodroski J, Wilson IA, Wyatt R, & Kwong PD (2005) Structure of a V3-containing HIV-1 gp120 core. *Science* 310(5750):1025-1028.

35. Guttman M, Cupo A, Julien JP, Sanders RW, Wilson IA, Moore JP, & Lee KK (2015) Antibody potency relates to the ability to recognize the closed, pre-fusion form of HIV Env. *Nat Commun* 6:6144.

36. Munro JB & Mothes W (2015) Structure and Dynamics of the Native HIV-1 Env Trimer. *J Virol* 89(11):5752-5755.

37. de Taeye SW, Ozorowski G, Torrents de la Pena A, Guttman M, Julien JP, van den Kerkhof TL, Burger JA, Pritchard LK, Pugach P, Yasmeen A, Crampton J, Hu J, Bontjer I, Torres JL, Arendt H, DeStefano J, Koff WC, Schuitemaker H, Eggink D, Berkout B, Dean H, LaBranche C, Crotty S, Crispin M, Montefiori DC, Klasse PJ, Lee KK, Moore JP, Wilson IA, Ward AB, & Sanders RW (2015) Immunogenicity of Stabilized HIV-1 Envelope Trimers with Reduced Exposure of Non-neutralizing Epitopes. *Cell* 163(7):1702-1715.

38. Julien JP, Lee JH, Ozorowski G, Hua Y, Torrents de la Pena A, de Taeye SW, Nieusma T, Cupo A, Yasmeen A, Golabek M, Pugach P, Klasse PJ, Moore JP, Sanders RW, Ward AB, & Wilson IA (2015) Design and structure of two HIV-1 clade C SOSIP.664 trimers that increase the arsenal of native-like Env immunogens. *Proc Natl Acad Sci U S A* 112(38):11947-11952.

39. Pugach P, Ozorowski G, Cupo A, Ringe R, Yasmeen A, de Val N, Derking R, Kim HJ, Korzun J, Golabek M, de Los Reyes K, Ketas TJ, Julien JP, Burton DR, Wilson IA, Sanders RW, Klasse PJ, Ward AB, & Moore JP (2015) A native-like SOSIP.664 trimer based on an HIV-1 subtype B env gene. *J Virol* 89(6):3380-3395.

40. Harris AK, Bartesaghi A, Milne JL, & Subramaniam S (2013) HIV-1 envelope glycoprotein trimers display open quaternary conformation when bound to the gp41 membrane-proximal external-region-directed broadly neutralizing antibody Z13e1. *J Virol* 87(12):7191-7196.

41. Diskin R, Marcovecchio PM, & Bjorkman PJ (2010) Structure of a clade C HIV-1 gp120 bound to CD4 and CD4-induced antibody reveals anti-CD4 polyreactivity. *Nat Struct Mol Biol* 17(5):608-613.

42. Mastronarde DN (2005) Automated electron microscope tomography using robust prediction of specimen movements. *J Struct Biol* 152(1):36-51.

43. Brilot AF, Chen JZ, Cheng A, Pan J, Harrison SC, Potter CS, Carragher B, Henderson R, & Grigorieff N (2012) Beam-induced motion of vitrified specimen on holey carbon film. *J Struct Biol* 177(3):630-637.

44. Campbell MG, Cheng A, Brilot AF, Moeller A, Lyumkis D, Veesler D, Pan J, Harrison SC, Potter CS, Carragher B, & Grigorieff N (2012) Movies of ice-embedded particles enhance resolution in electron cryo-microscopy. *Structure* 20(11):1823-1828.

45. Grant T & Grigorieff N (2015) Measuring the optimal exposure for single particle cryo-EM using a 2.6 Å reconstruction of rotavirus VP6. *Elife* 4:e06980.

46. Tang G, Peng L, Baldwin PR, Mann DS, Jiang W, Rees I, & Ludtke SJ (2007) EMAN2: an extensible image processing suite for electron microscopy. *J Struct Biol* 157(1):38-46.

47. Rohou A & Grigorieff N (2015) CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J Struct Biol* 192(2):216-221.

48. Scheres SH (2012) RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J Struct Biol* 180(3):519-530.
49. Goddard TD, Huang CC, & Ferrin TE (2007) Visualizing density maps with UCSF Chimera. *J Struct Biol* 157(1):281-287.
50. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, & Zwart PH (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):213-221.
51. Zhang Y & Skolnick J (2005) TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res* 33(7):2302-2309.
52. Siciliano B & Khatib O (2008) Springer handbook of robotics (Springer, Berlin) pp Ix, 1611 p.
53. West AP, Jr., Scharf L, Horwitz J, Klein F, Nussenzweig MC, & Bjorkman PJ (2013) Computational analysis of anti-HIV-1 antibody neutralization panel data to identify potential functional epitope residues. *Proc Natl Acad Sci U S A* 110(26):10598-10603.

Figure Legends.

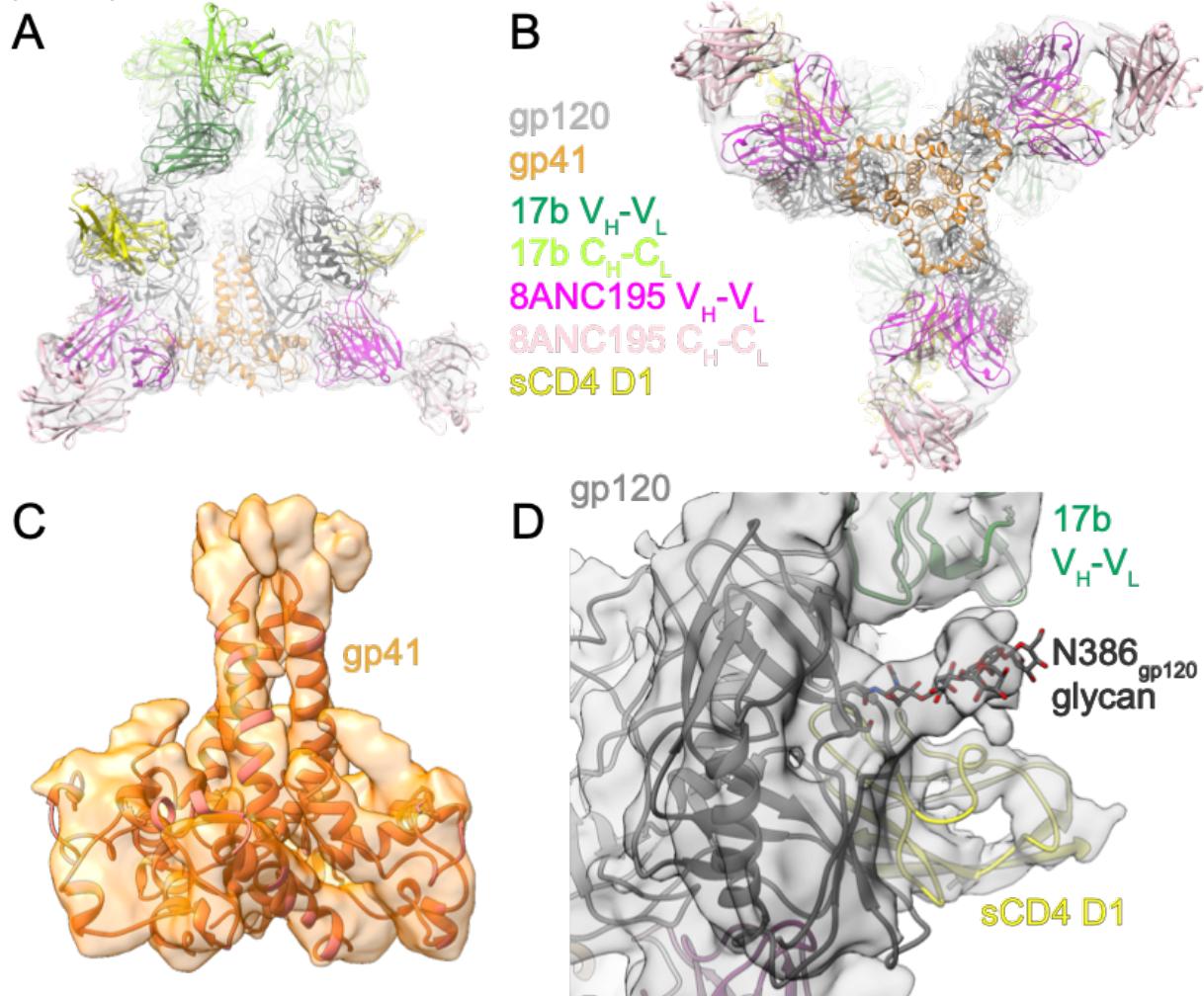


Fig. 1. 8.9 Å EM reconstruction of Env-sCD4-17b-8ANC complex. (A, B) Electron density fit by coordinates for gp120 (gray), gp41 (orange), sCD4 D1 (yellow), 17b  $V_H$ - $V_L$  (forest green), 17b  $C_H$ - $C_L$  (chartreuse), 8ANC195  $V_H$ - $V_L$  (magenta), and 8ANC195  $C_H$ - $C_L$  (pink). N-linked glycan coordinates are shown as sticks. (A) Side view in which the three-fold symmetry axis of the BG505 trimer is vertical. (B) Bottom view looking down the three-fold symmetry axis of the BG505 trimer. (C) Side view of density of gp41 portion of the BG505 trimer. (D) Close-up of density near the N-linked glycan attached to BG505 Asn368<sub>gp120</sub>.

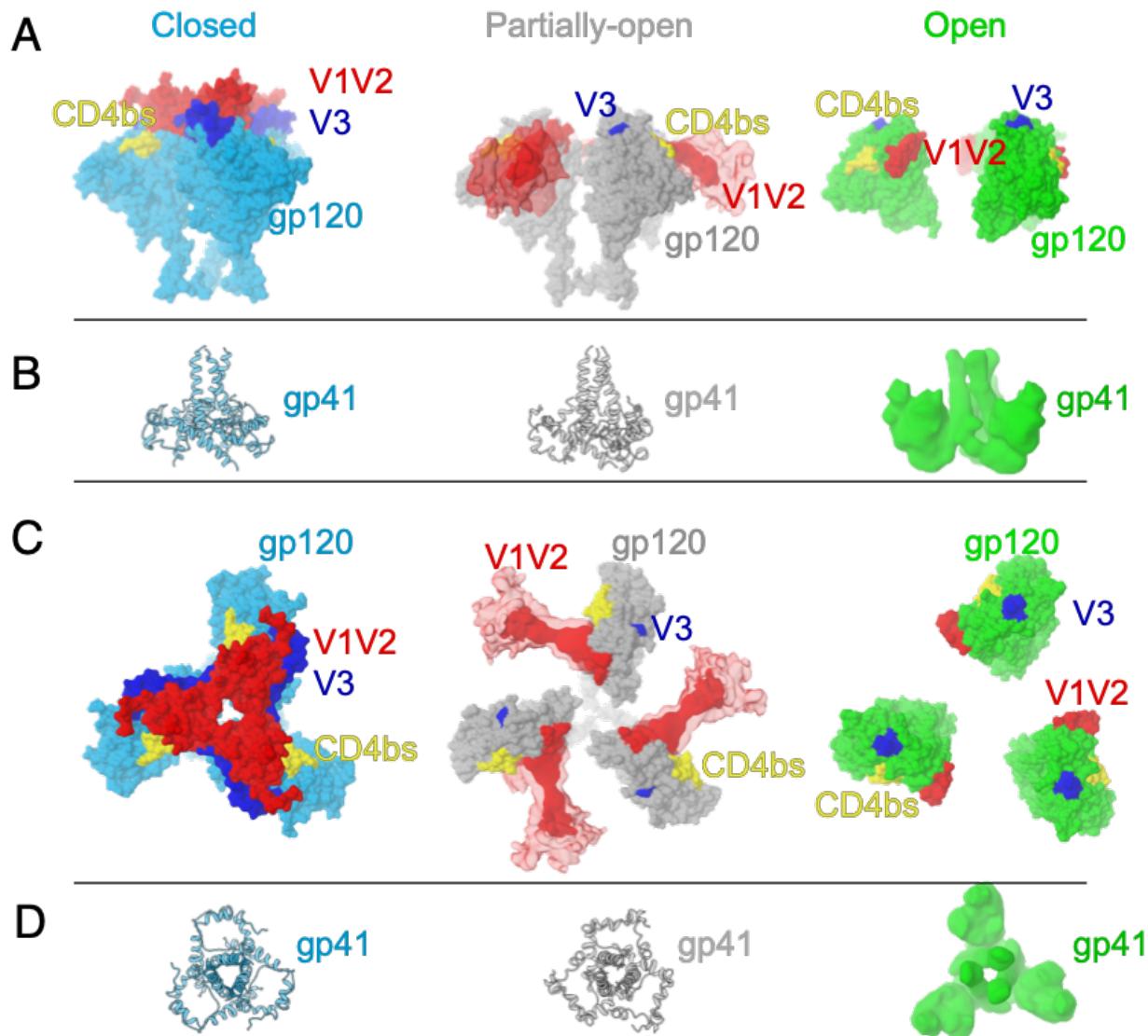


Fig. 2. CD4-induced structural changes in Env trimers. (A) gp120 surface representations for the closed (PDB 5T3X) (blue), partially-open (this study) (grey), and sCD4-bound open Env structures (PDB 3DNO) (green) as seen from the side. V1V2 loops (red) are depicted as surface representations for the closed and open structures and as EM density for the partially-open structure. Locations of V3 (blue) and the CD4 binding site (CD4bs) (yellow) are depicted as surface representations. (B) gp41 in ribbon representation (closed and partially-open Env structures) or as density (open Env structure) for Env structures. As gp41 coordinates for an open Env structure were unavailable, we used the density from the single particle EM structure of an open KNH1144-17b complex (EMDB 5462) (7) (C) Top view of gp120 representations shown in panel A. (C) Top view of gp41 representations shown in panel A.

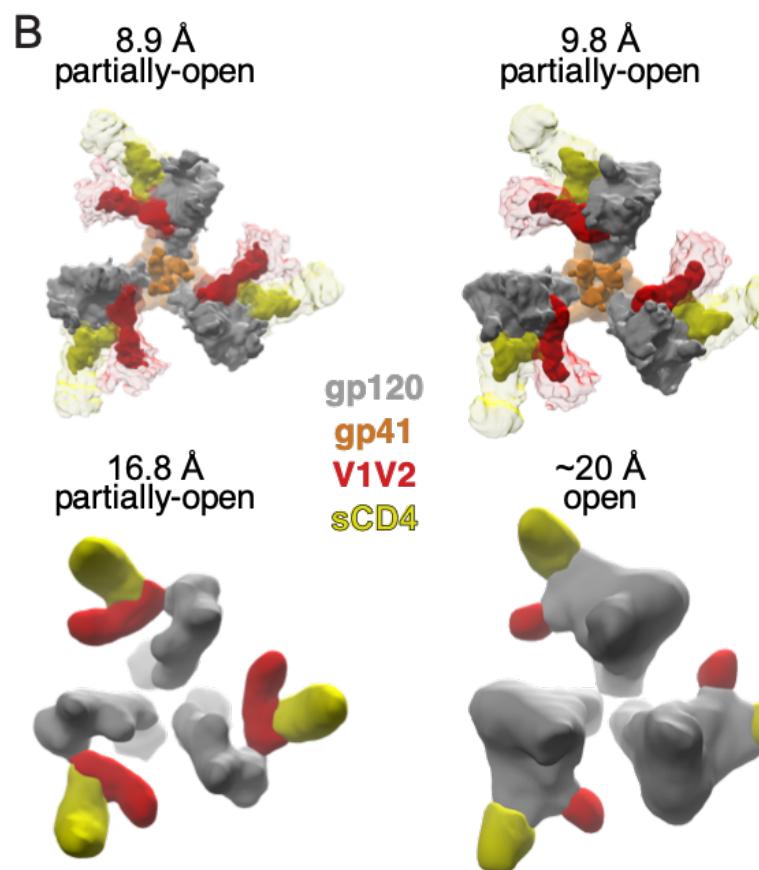
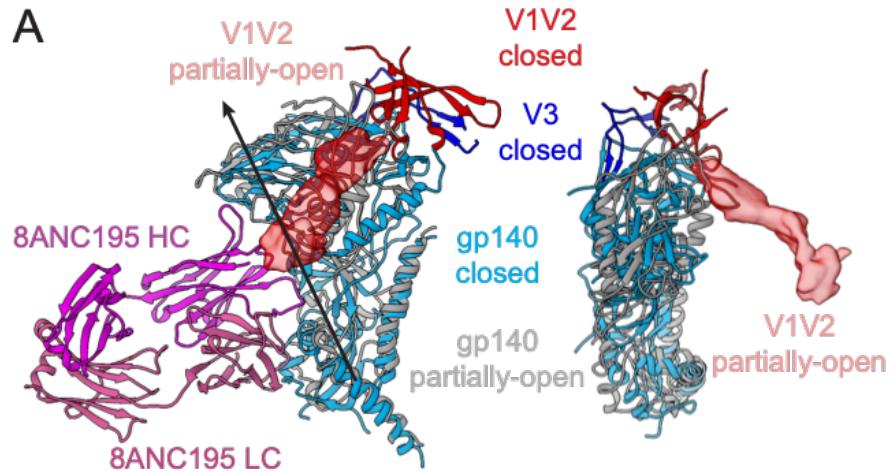


Fig. 3. Putative V1V2 densities. (A) Superposition of the gp140 (gp120 plus gp41) coordinates from one monomer in a closed BG505 Env trimer structure (PDB 5T3X) (blue) and the gp140 coordinates from the 8.9 Å partially-open BG505 Env trimer reported here (gray) as seen from two different orientations. The gp140 monomers were aligned using the 3-fold symmetry axis of the gp41 trimer. The arrow in the left panel marks the axis about which a rotation of 30° and a translation of 2.6 Å relates the gp120 in the closed structure to the gp120 in the partially-open structure. The V1V2 loop (red) is depicted as a ribbon for the closed trimer and as EM density for the partially-open structure. The V3 loop (dark blue) is depicted as a ribbon for closed trimer and not shown in the partially-open structure because it was disordered. The 8ANC195 Fab (magenta heavy chain; light pink light chain) is shown in the left panel based on its position with respect to the partially-open gp140, illustrating that its epitope at the gp120-gp41 interface does not undergo extensive changes. (B) sCD4-proximal densities (red) in four independent CD4-bound Env structures: the 8.9 Å and 9.8 Å cryo-EM reconstructions of the Env-sCD4-17b-8ANC complex (this study), the 16.8 Å negative stain reconstruction of the Env-sCD4-17b-8ANC complex (18), and the ~20 Å reconstruction of a sCD4-Env trimer structure derived from sub-tomogram averaging of virion-bound Env spikes (4). High contour level densities in the Env-sCD4-17b-8ANC reconstructions are shown in bright red and bright yellow for V1V2 and sCD4, respectively, with lower contour level densities in lighter colors.

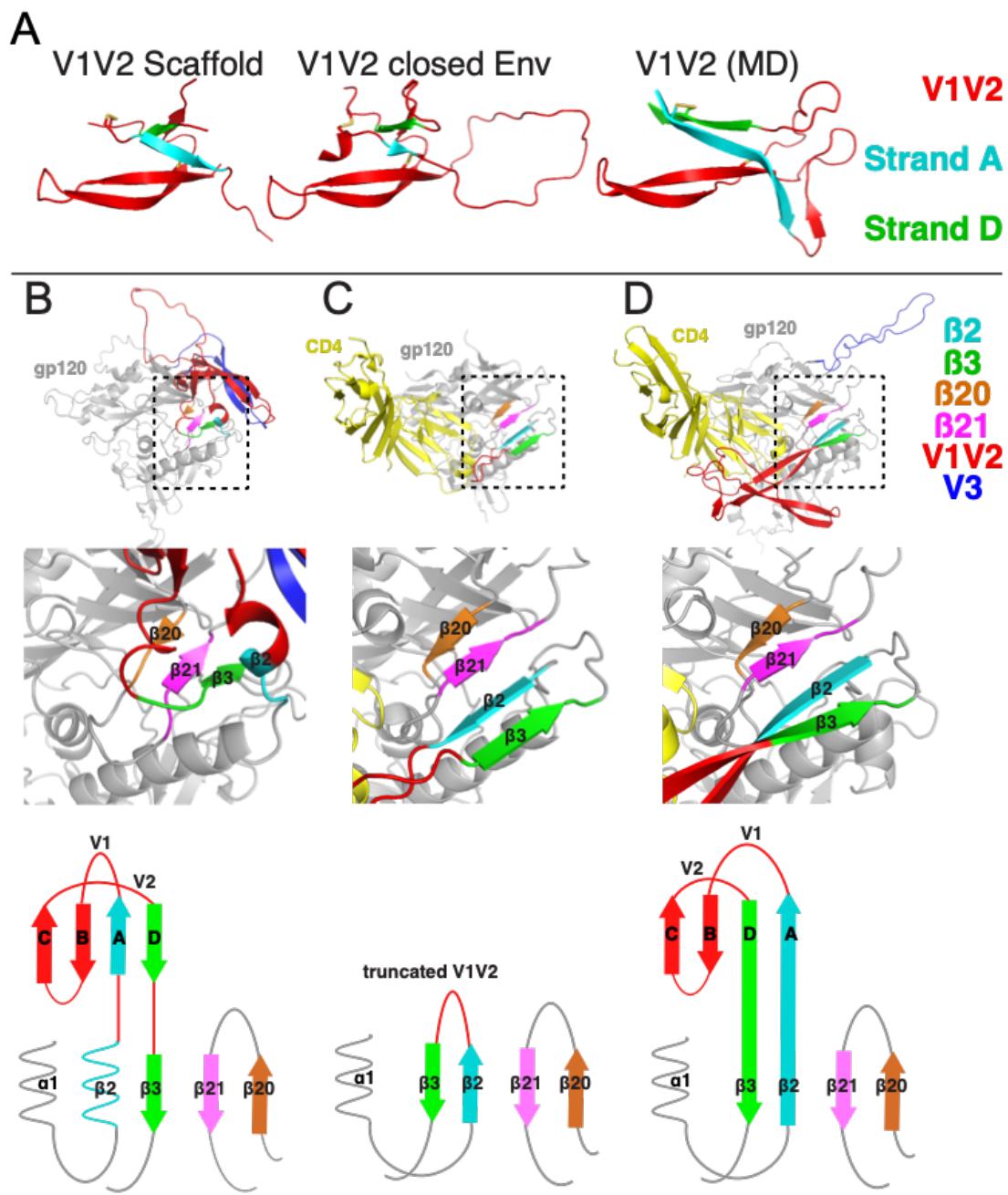


Fig. 4. V1V2 loop structures. (A) V1V2 folding topologies in V1V2 scaffold (PDB 5ESV) (32, 33), closed BG505 trimer (PDB 5FYJ) (10, 19), and molecular dynamics model of repositioned V1V2 in full-length CD4-bound gp120 (21).  $\beta$ -strand nomenclature in V1V2 is the same as in (32). Disulfide bonds are shown as yellow sticks. (B-D) gp120s from structures of closed BG505 trimer (PDB 5FYJ) (19) (B), sCD4-bound monomeric gp120 core (truncated V1V2 and V3 loops) (PDB 1RZJ) (30) (C),

and the molecular dynamics model of full-length CD4-bound gp120 (21) (D). Top panels show structural overviews. Middle panels show close-up views of the regions in the boxed areas. Bottom panels show topology diagrams of the bridging sheet (adapted from (16)). V1V2 is red and V3 is blue. Strands  $\beta$ 2 and  $\beta$ 3, which precede and follow V1V2 in the gp120 sequence, are cyan and green, respectively. gp120 strands  $\beta$ 20 and  $\beta$ 21, which form a  $\beta$ -sheet with  $\beta$ 2 and  $\beta$ 3 in sCD4-bound gp120 structures (31) are orange and magenta, respectively.

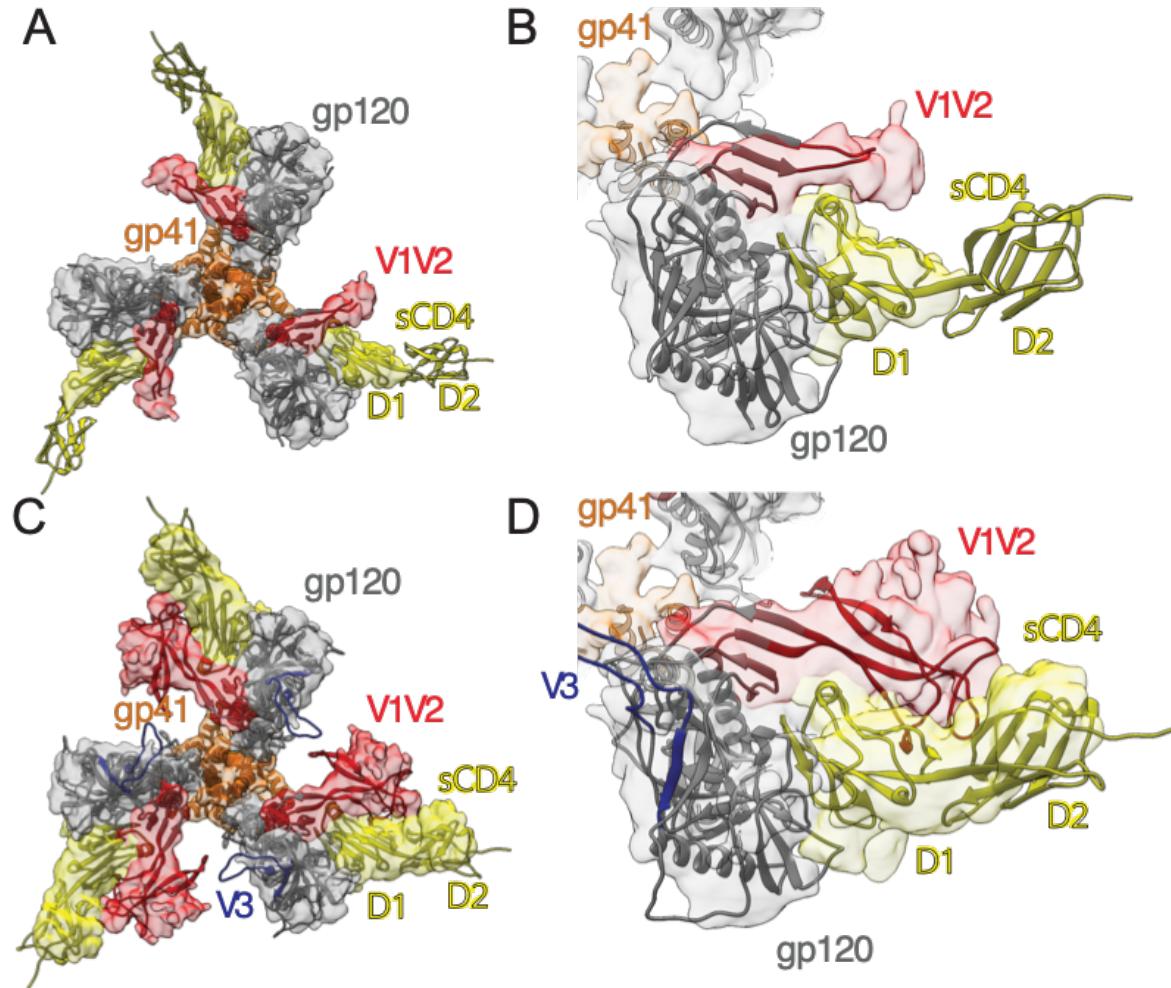


Fig. 5. V1V2 fitting to EM density. Densities are gray (gp120), orange (gp41), yellow (sCD4), and red (V1V2). (A) EM density map (high contour level) fit with coordinates of a sCD4-bound monomeric gp120 core (truncated V1V2 and V3 loops) (PDB 1RZJ) (30). (B) Close-up of map and coordinates in panel A showing putative contacts between V1V2 loop stem and CD4 D1. (C) EM density map (low contour level for V1V2 and sCD4 regions) fit with coordinates of the molecular dynamics model of full-length CD4-bound gp120 (21). The V3 loop does not fit in the density. (D) Close-up of map and coordinates in panel C.

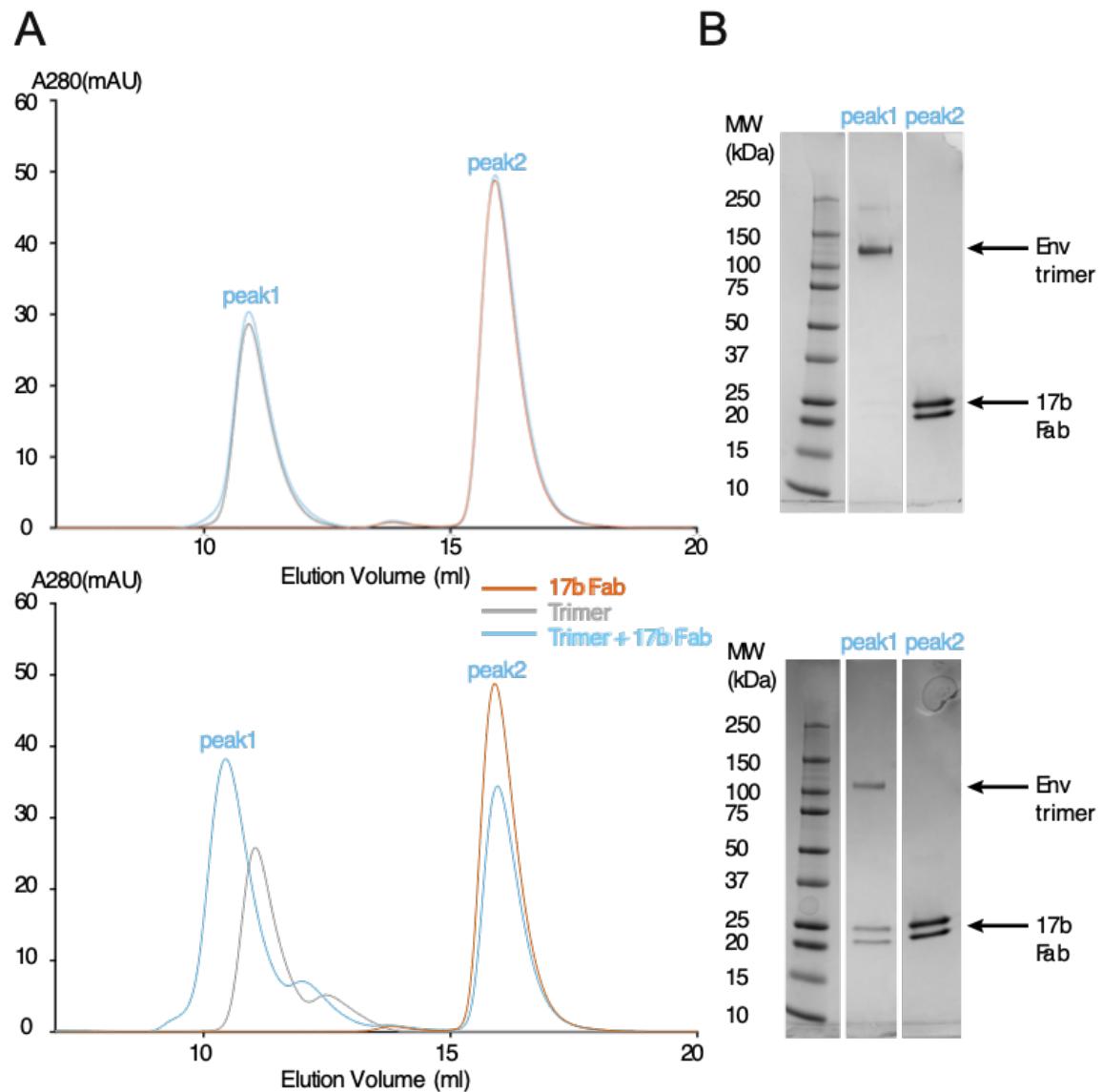


Fig. 6. 17b interactions with BG505 and BG505- $\Delta$ V1V2. (A) SEC profile demonstrating that BG505- $\Delta$ V1V2, but not BG505, binds 17b Fab. (B) SDS-PAGE analysis of SEC fractions.

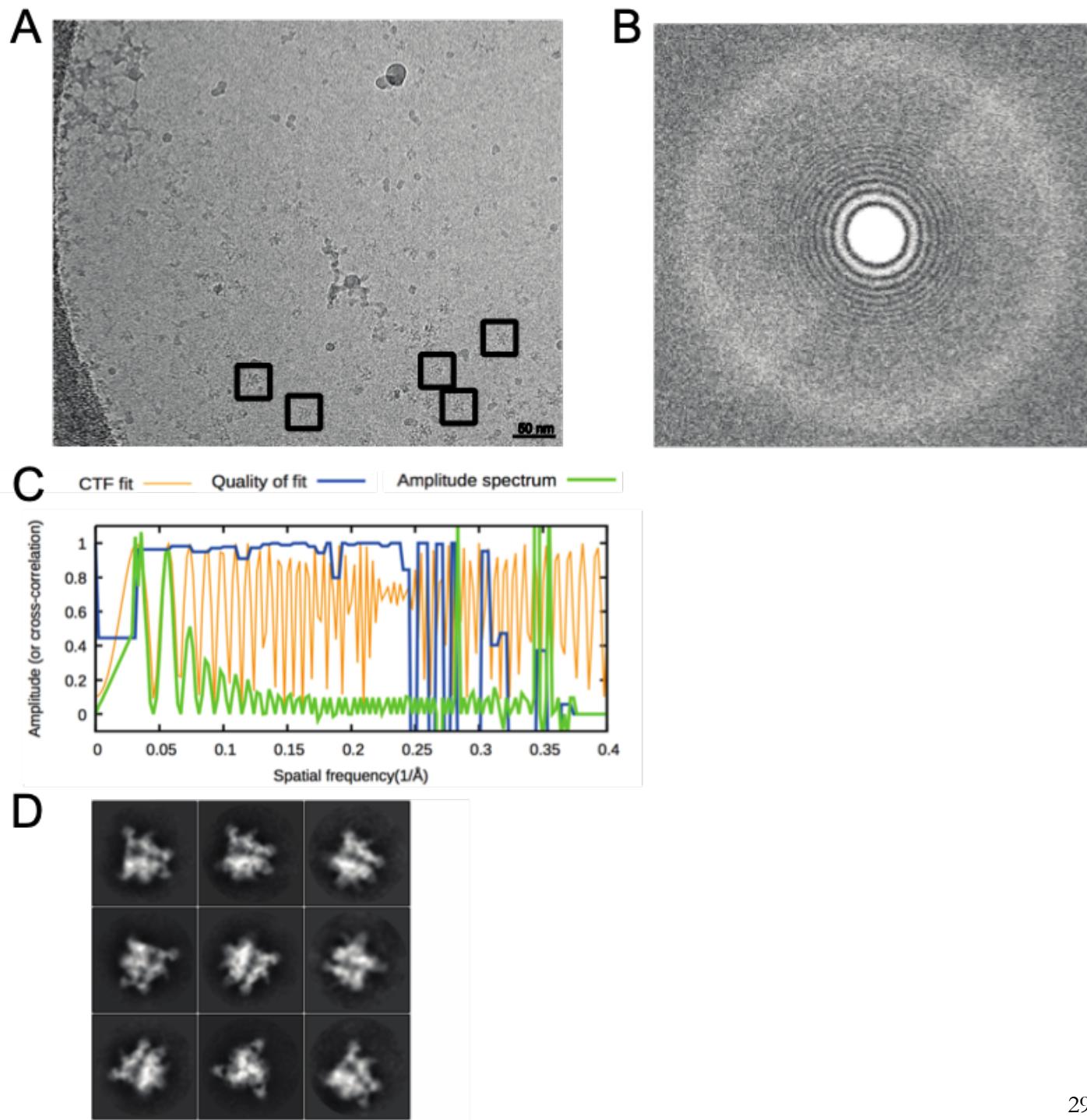


Fig. S1. 8.9 Å EM structure determination. (A) Image of the sample grid used for 8.9 Å reconstructions of the Env-sCD4-17b-8ANC complex. Examples of complex particles are boxed. The accumulated dose and defocus values were  $\sim 25$  e/ $\text{\AA}^2$  and  $\sim 2.2$  μm underfocus. (B) Fourier transform of the micrograph in panel A. (C) CTF fitting of the left micrograph (generated using CTFFIND4) shows a good fit to 4.0 Å (D) Representative reference-free 2D class averages calculated by RELION.

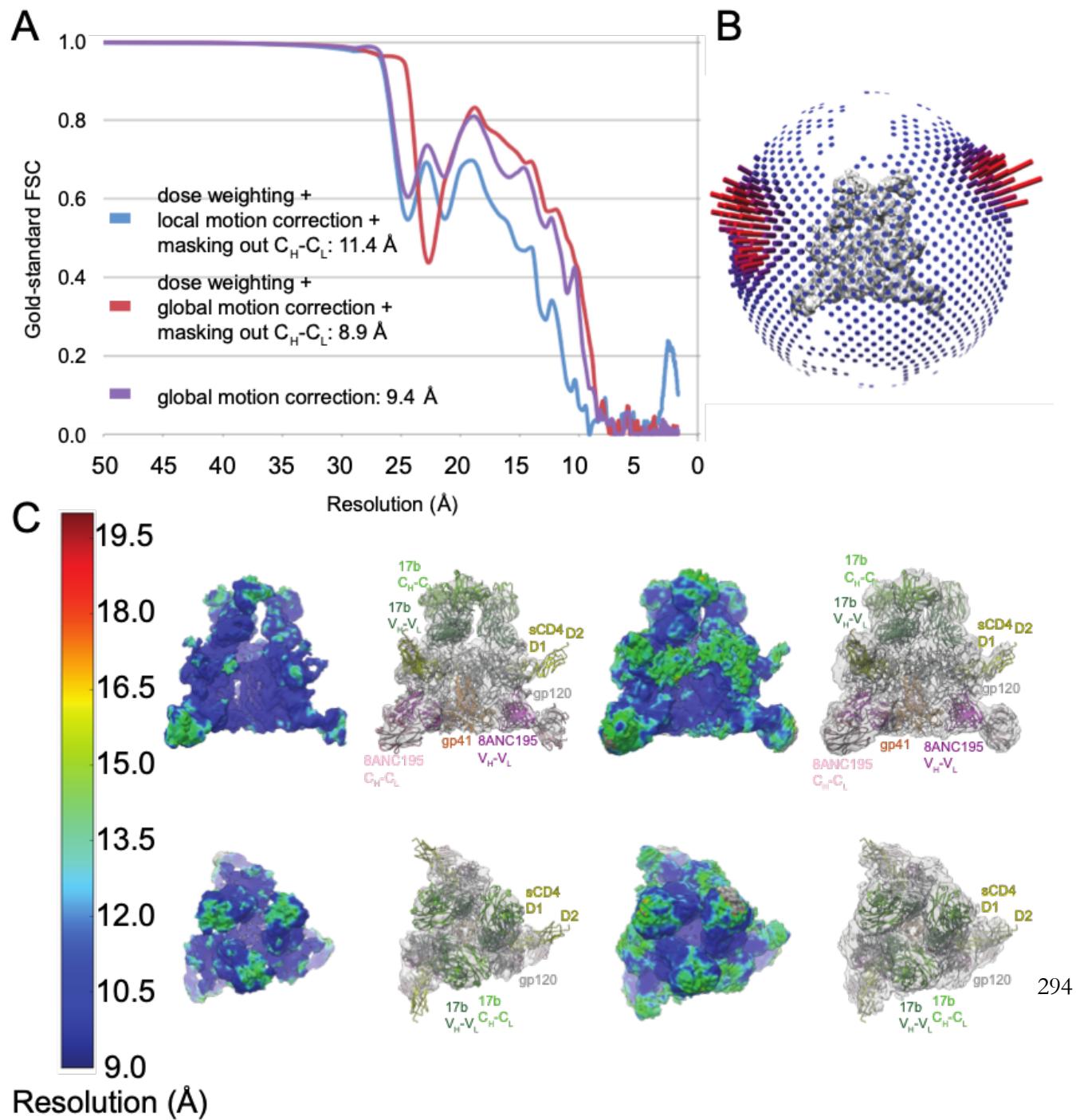


Fig. S2. EM structure analysis. (A) Gold-standard FSCs of reconstructions generated using different strategies. The 8.9 Å reconstruction was generated using dose weighting and masking out the Fab  $C_H$ - $C_L$  domains. (B) Orientation distribution of the 8.9 Å reconstruction. (C) Local resolution estimation in the 8.9 Å reconstruction. The Env-sCD4-17b-8ANC complex is shown as side (top row) and top (bottom row) views. Local resolution estimations are shown for a high contour level (left) and a lower contour level (right) with fitted coordinates shown beside each. In the local resolution map, gray represents resolutions below 20 Å. The lower contour level was necessary to visualize density for the sCD4 D2 domain and V1V2 loop. These regions were estimated to be ~15 Å resolution.

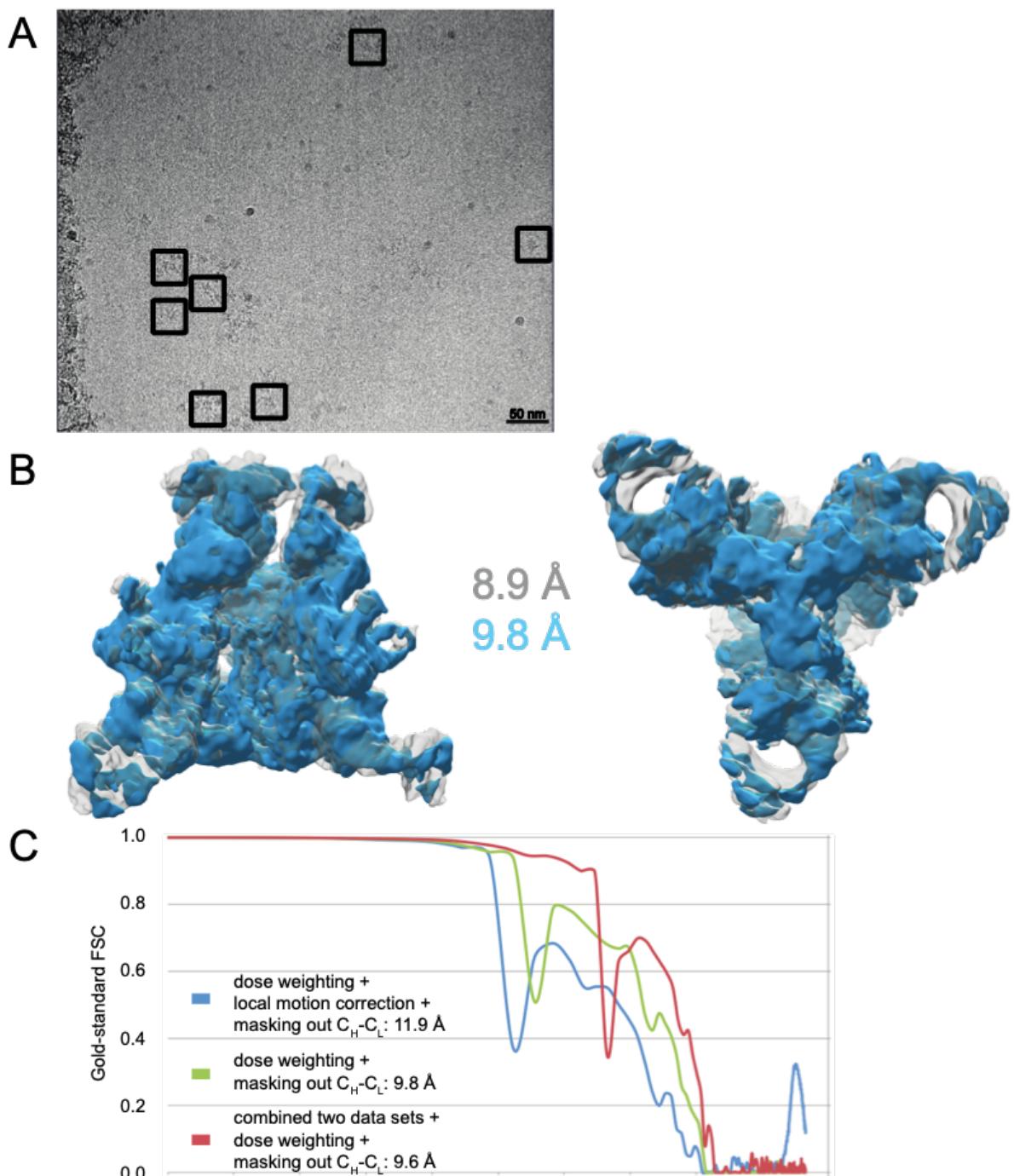


Fig. S3. Comparison of 8.9 Å and 9.8 Å Env-sCD4-17b-8ANC reconstructions. (A) Image of sample grid used for 9.8 Å reconstruction of the Env-sCD4-17b-8ANC complex. Examples of complex particles are boxed. The accumulated dose and defocus values were ~68 e-Å<sup>2</sup> and ~3.3 μm defocus. (B) Superimposition of 8.9 Å and 9.8 Å reconstructions seen from the side (left) and bottom (right). (C) Gold-standard FSCs of reconstructions generated using different strategies. The 9.8 Å reconstruction was generated using dose weighting and masking out the Fab C<sub>H</sub>-C<sub>L</sub> domains. The calculated resolution was 9.6 Å when combining the two independent data sets (see Methods), using dose weighting, and masking out the Fab C<sub>H</sub>-C<sub>L</sub> domains.

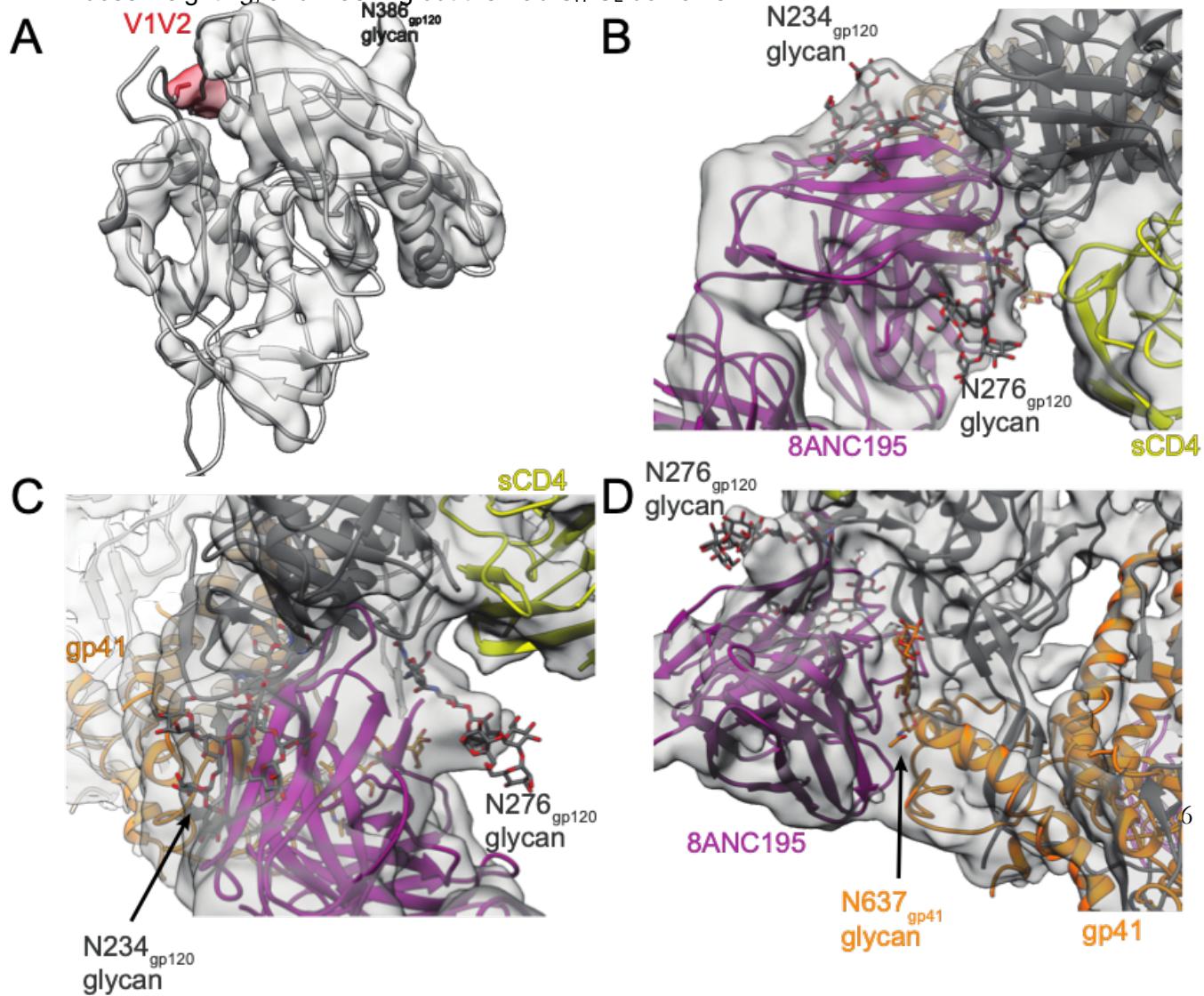


Fig. S4. Electron density in the 8.9 Å EM reconstruction. Close-up views of density in gp120 (A) and near the N-linked glycans attached to BG505 Asn234gp120 (B), BG505 Asn276gp120 (C), and BG505 Asn637gp41 (D). Glycans are shown as sticks.

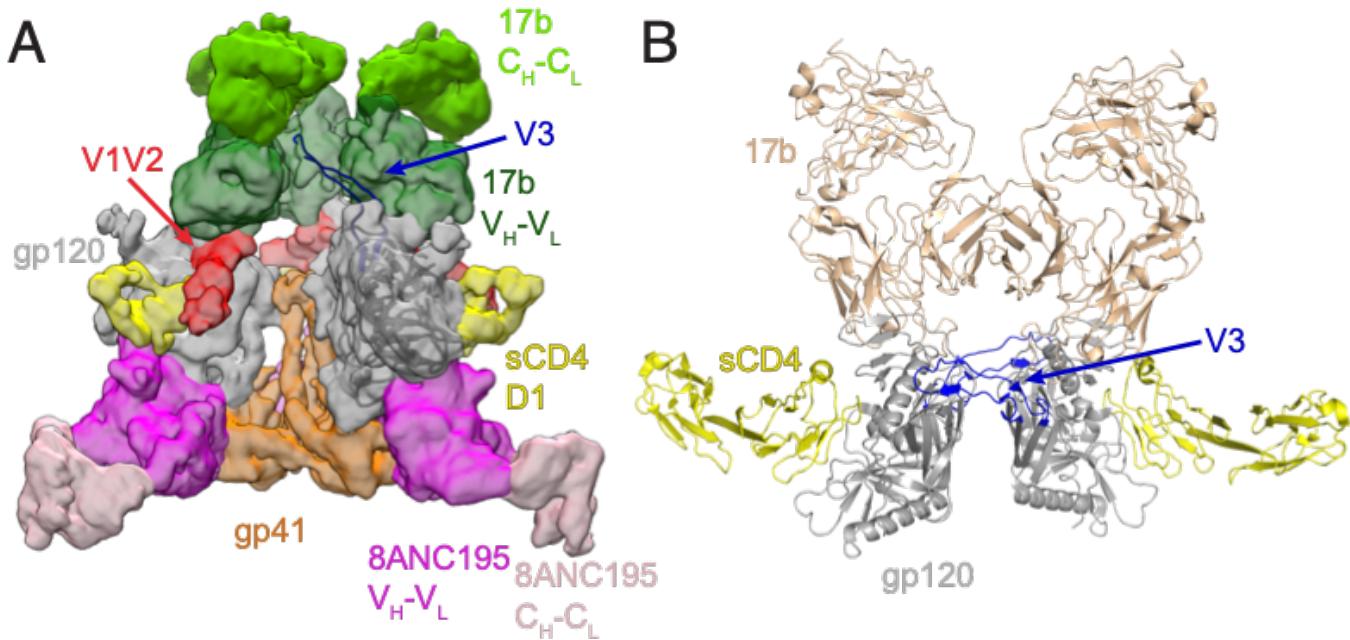


Fig. S5. V3 loop. (A) The gp120 V3 could not be localized in EM density of the sCD4-bound partially-open BG505 complex. The V3 loop position in the coordinates shown is derived from its position in a sCD4-bound gp120 core structure with an ordered V3 loop (34) (PDB 2QAD). (B) Crystal contacts in PDB 2QAD, a complex between a gp120 core (gray) with an ordered V3 loop (blue), sCD4 (yellow), and a CD4i Fab (wheat). The ordered V3 loops contact adjacent gp120s in the crystal.

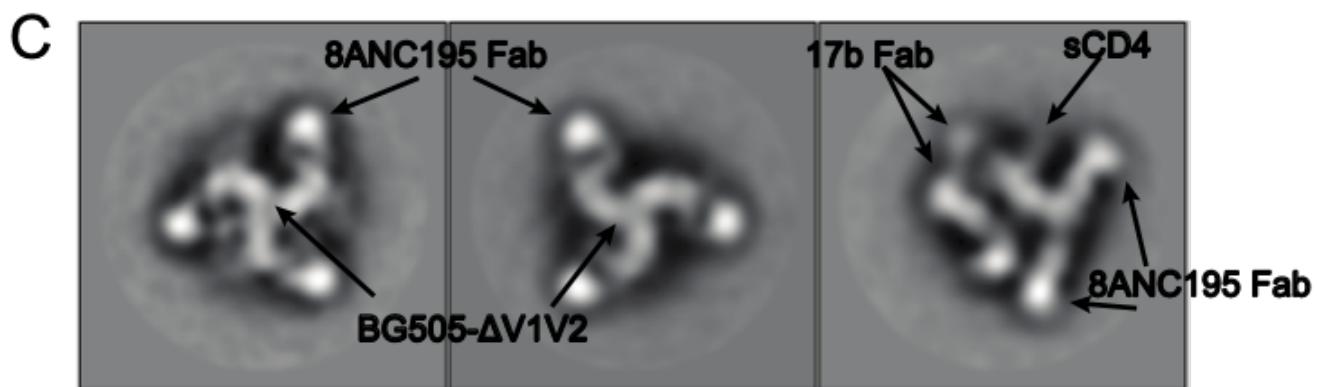
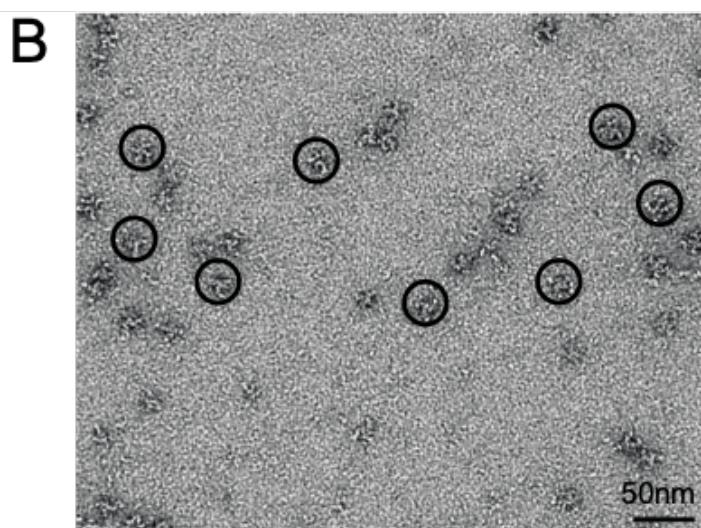
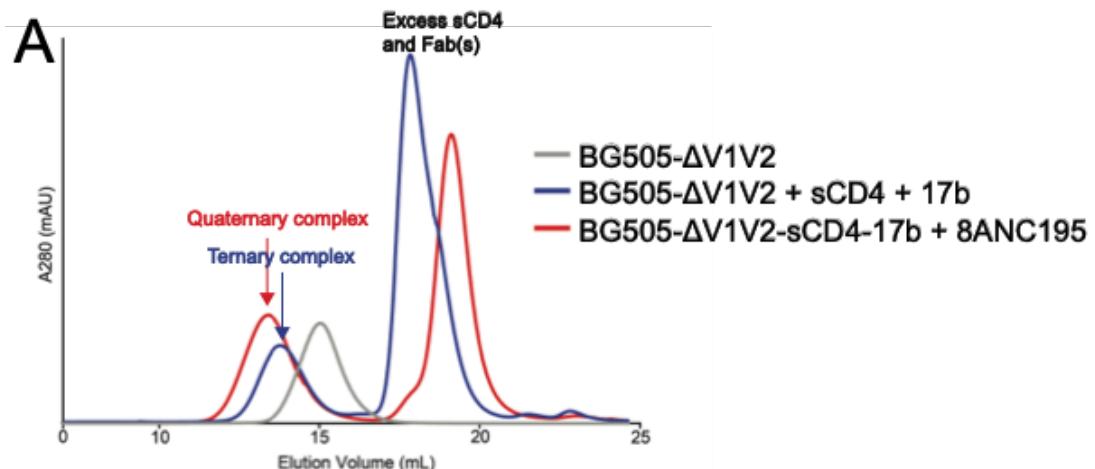


Fig. S6. Characterization of BG505-ΔV1V2. (A) Size exclusion chromatography (SEC) profiles showing SEC purification steps. The BG505-ΔV1V2-sCD4-17b ternary complex (blue) was purified first and then 8ANC195 Fab was added to form the quaternary complex (red). The faster-migrating peaks (12-15 mL) correspond to BG505-ΔV1V2 alone (gray), the ternary BG505-ΔV1V2-sCD4-17b complex (blue), and the BG505-ΔV1V2-sCD4-17b-8ANC195 complex (red). The slower migrating peaks (17-21 mL) correspond to free sCD4 and Fabs. (B) Image of the sample grid used for negative stain single particle reconstruction. Examples of complex particles are circled. (C) 2D class averages of the BG505-ΔV1V2-sCD4-17b-8ANC195 complex. Arrows point to identified ligands.