Increasing the breadth and potency of a neutralising single domain antibody against the influenza haemagglutinin stem

Andrew David Tung Yep

Thesis submitted for the degree of
Doctor of Philosophy

Division of Infection and Immunity
University College London

2022

I, Andrew David Tung Yep, confirm that the work presented in my thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

Among influenza A group 1 viruses, those with the haemagglutinin (HA) subtypes H1, H2, H5 and H9 are considered to pose the greatest pandemic risk. The alpaca-derived single domain antibody R1a-B6 binds to a conserved epitope on the influenza HA stem, resulting in neutralisation by membrane fusion inhibition. R1a-B6 has demonstrated *in vivo* efficacy against H1 and H5 viruses but lacks potency against H2 and H9 viruses, tested *in vitro*.

We created a rationally designed library of millions of R1a-B6 variants displayed on the surface of yeast by incorporating amino acid variation from clonally related antibodies. These cells were sorted by flow cytometry in successive rounds according to their binding to H9 HA, resulting in enrichment of high-affinity H9 binding variants.

R1a-B6 variant WC21 demonstrated dramatic improvements to affinity, ELISA EC50 and pseudovirus neutralisation titre against H9 and H2 antigens and maintained effectiveness against H1 and H5 antigens. The improvements were mediated by one CDR1 substitution (R31I) and two CDR3 substitutions (D95N and Y102S). R1a-B6 single substitution variants R31I, D95N and Y102S each slowed dissociation from H2 and H9 antigens and these three substitutions in combination demonstrated the same affinity improvements as WC21.

We used saturation mutagenesis and deep mutational scanning to identify HA substitutions which led to loss of antibody binding. We demonstrated that WC21 has a more compact epitope footprint than R1a-B6 and speculate that this may represent a higher genetic barrier for viral escape from WC21. Furthermore, this experiment highlighted an H2-specific HA polymorphism, HA2 I45F, which abrogated R1a-B6 binding but not WC21 binding. Overcoming the barrier to binding posed by HA2 I45F is suggested to be a primary reason for the dramatic increase in H2 potency. Our approach informs strategies which may be broadly applicable to the molecular evolution of monoclonal antibodies against antigenically variable pathogens.

Impact statement

Advances in the isolation and characterisation of monoclonal antibodies have facilitated the discovery of antibodies able to neutralise a diverse range of influenza viruses. A large amount of current influenza research aims to develop universal influenza vaccines capable of eliciting these broadly neutralising antibodies to promote influenza immunity to a wide range of strains of the virus. Considerable advances have been made in this field, but an approved and effective universal influenza vaccine will not be available for several years. Moreover, some of the most vulnerable populations such as the elderly and the immunocompromised are unable to mount strong immune responses so may receive little protection from a vaccine.

An alternative and complementary approach to reducing the impact of seasonal influenza and improving preparedness against influenza pandemics is the use of anti-influenza antibodies as drugs and diagnostic reagents. R1a-B6 is one such broadly neutralising anti-influenza antibody derived from a camelid (specifically an alpaca). Camelid derived antibodies have a lower molecular mass and are more modular and more cost effective to produce than human antibodies. We isolated and characterised a variant of R1a-B6 which can bind to and neutralise a wider range of influenza viruses than the parental antibody. The modified antibody, WC21, has dramatically improved potency against two key pandemic risk lineages of influenza. We demonstrated that the binding site of WC21 is more compact than that of R1a-B6 and that WC21 will bind to this site despite sequence changes at key amino acids.

The wide range of influenza viruses neutralised potently by WC21 and the flexible formatting of camelid antibodies, makes WC21 an attractive candidate for several anti-influenza applications. For example, virus vectored gene therapy could induce constant expression of WC21 in patients, providing long term protection from many influenza lineages independent of a functional immune system.

Thus far, few studies have been published on broadening the neutralisation range of a monoclonal antibody. The approach we took to optimising R1a-B6 combined bioinformatic data with experimental data to decide the location of mutations to insert into the R1a-B6 gene. The techniques used in this thesis can

be applied to the development of other antibodies directed at antigenically diverse viruses such as HIV, hepatitis C virus and SARS-CoV2, potentially reducing the burden of these diseases.

Acknowledgements

I'd like to extend my thanks to the many people who have taught, advised and encouraged me throughout the 4 years I've spent on this project. I've learnt, not only many new experimental techniques, but also skills in experimental design, literature review and academic writing. None of this would have been possible without the generous support of so many people.

My main supervisors, Dr Simon Hufton and Dr Yasuhiro Takeuchi, have been constant sources of advice, mentorship and feedback on experiments, presentations and writing. They have spent countless hours with me discussing all aspect of this project. I could not have asked for a more supportive supervisory team.

The members of my thesis committee Dr Othmar Engelhardt and Dr Joe Grove have provided critical advice and feedback on this project.

Ryan Mate, Martin Gordon and Walter Ramage carried out the Illumina sequencing in this project and advised on the data analysis.

Thank you to Dr Anna Nowocin, Rachel Peraj and Dr Sandra Diebold for all the help with flow cytometry and cell sorting.

Chris Ball helped immensely with protein expression and purification.

Thank you to Dr Giada Mattiuzzo, Dr Emma Bentley, Stephanie Routley, Dr Joanne Del Rosario and Professor Nigel Temperton for teaching me to make and use pseudoviruses and providing the pseudoviruses used in this project.

Finally, I'd like to thank my family and friends, especially my parents, for their consistent interest, encouragement and support.

Abbreviations

2x YT	2x yeast tryptophan
A450	Absorption of 450nm wavelength light
AAV	Adeno-associated virus
ADCC	Antibody dependent cellular cytotoxicity
ADE	Antibody dependent enhancement
AF405	Alexa fluor 405
AF488	Alexa fluor 488
AF647	Alexa fluor 647
BLAST	Basic local alignment search tool
BSA	Bovine serum albumin Copy DNA (DNA produced by reverse transcription from RNA)
CDR	Complementarity determining regions
cRNP	Copy ribonucleoprotein
Cryo-EM	Cryogenic electron microscopy
DMSO	Dimethyl sulfoxide
E.coli	*Escherichia coli*
EC50	50% saturation concentration in an ELISA
EDTA	Ethylenediaminetetraacetic acid
ELISA	Enzyme-linked immunosorbent assay
epPCR	Error prone PCR
FACS	Fluorescence activated cell sorting
FR	Framework region
HA	Haemagglutinin
HAI	Haemagglutination inhibition
HEK293	Human embryonic kidney 293
HRP	Horseradish peroxidase
IC50	50% inhibitory concentration
IPTG	Isopropyl β-D-1-thiogalactopyranoside
M1	Matrix protein 1
M2	Matrix protein 2

MFI	Mean fluorescence intensity
NA	Neuraminidase
NP	Nucleoprotein
NS1	Non-structural protein 1
NS2	Non-structural protein 2
OD600	Optical density to 600nm wavelength light
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
Periprep	Periplasm preparation
Phage	Bacteriophage
RBS	Receptor binding site
RLU	Relative luminescence units
SA	Sialic acid
SCK	Single cycle kinetics
SD/CAA	Synthetic defined medium with casamino acids
SDS	Sodium dodecyl sulphate
SG/R+CAA	Synthetic defined galactose and raffinose medium with casamino acids
SOC	Super optimal broth with catabolite repression
soePCR	Splicing by overlap extension PCR
SPR	Surface plasmon resonance
TMPRSS-2/4	Transmembrane protease serine S1 member 2/4
vRNP	Viral ribonucleoprotein
WT	Wild type
YPD	Yeast extract peptone dextrose

Contents

1. Introduction

The influenza virus is one of the earliest known pathogens to affect humanity. Past influenza pandemics have been described in detail for hundreds of years and sparser records dating back thousands of years appear to describe influenza outbreaks and symptoms[1, 2]. Currently, influenza causes hundreds of millions of infections and hundreds of thousands of deaths every year in winter outbreaks of so-called seasonal flu[3, 4]. Seasonal flu tends to be caused by viruses closely related genetically and antigenically to seasonal flu strains in previous years, meaning that the population has a degree of immunity to the virus. Therefore, the severity of disease is often ameliorated[5]. Influenza viruses which are antigenically novel to the human population tend to originate from animal reservoirs, either from direct zoonotic transmission or through reassortment events with human viruses. An immunologically naïve population can allow a novel strain of influenza to transmit more effectively than seasonal flu and, in some cases, cause more severe disease. A global outbreak of a novel influenza strain is referred to as pandemic flu and, following a pandemic, these novel influenza strains are usually subsumed into the mixture of circulating viruses making up seasonal flu[6]. The emergence of SARS-CoV2 has reiterated the health, social and economic cost of a highly infectious respiratory pandemic in an era of commonplace worldwide travel[7]. Therefore, the value of protecting the population as a whole and vulnerable populations in particular, is immeasurable.

1.1. Influenza virus biology and life cycle

Influenza is an enveloped virus of the family *Orthomyxoviridae* with a segmented negative sense RNA genome. Known influenza virus genera are *Alphainfluenzavirus*, *Betainfluenzavirus*, *Gammainfluenzavirus* and *Deltainfluenzavirus* (influenza A, B, C and D)[8, 9]. Influenza C and D are closely related to each other and cause only mild disease in humans. Influenza C circulates in humans and pigs, predominantly infecting children and inducing long lasting influenza C immunity. Influenza D appears to be primarily a bovine virus but transmission to humans is inferred from serological data[10].

Influenza A and B currently pose a much greater risk to human health. Viruses of these two genera are the causative agents of yearly seasonal flu outbreaks and are capable of both more severe disease than influenza C and D as well as greater rates of antigenic evolution. Year on year antigenic evolution (or drift) facilitates repeated infections of individuals from the same lineage of virus[11].

Influenza B is primarily a human virus and sustained transmission in other animals has not been observed. Consequently, current theories propose that the antigenic novelty required for a new pandemic is unlikely to be produced in an influenza B virus due to the lack of animal reservoir[12].

The primary hosts of influenza A are wild waterfowl such as ducks, geese and gulls in which influenza is a gastrointestinal and respiratory pathogen transmitted through the faecal-oral route. However, influenza A is also able to infect and transmit among humans and a wide variety of livestock and companion animals such as poultry, pigs, dogs and horses. In mammals, influenza A is a respiratory pathogen[13].

Co-infection of cells with two different lineages of influenza A virus can result in a reassortment event in which hybrid viruses are produced with one or more segment from each parent virus. Reassortment events within and transmission events between livestock animals and then into the human population appears to be a primary mechanism by which influenza A pandemics start[14].

Influenza A and B viruses are structurally alike and share a similar cellular life cycle. A virion schematic is shown in figure INT 1a and a simplified influenza life cycle is shown in figure INT 1b.

Viral envelopes are populated with 3 viral proteins, haemagglutinin (HA), neuraminidase (NA) and matrix protein 2 (M2). HA, in its immature form (HA0), is a homotrimer. When each monomer has been proteolytically cleaved into 2 subunits (called HA1 and HA2), the HA protein is mature. NA is a homotetramer and M2 is a monomeric transmembrane proton channel. Lining the inside of the envelope is matrix protein 1 (M1) and at the centre of the virion is the 8-segment genome. Each genome segment is packaged in a viral ribonucleoprotein complex

(vRNP) consisting of the RNA segment, nucleoprotein (NP) and RNA polymerase (subunits PB1, PB2 and PA) [15, 16].

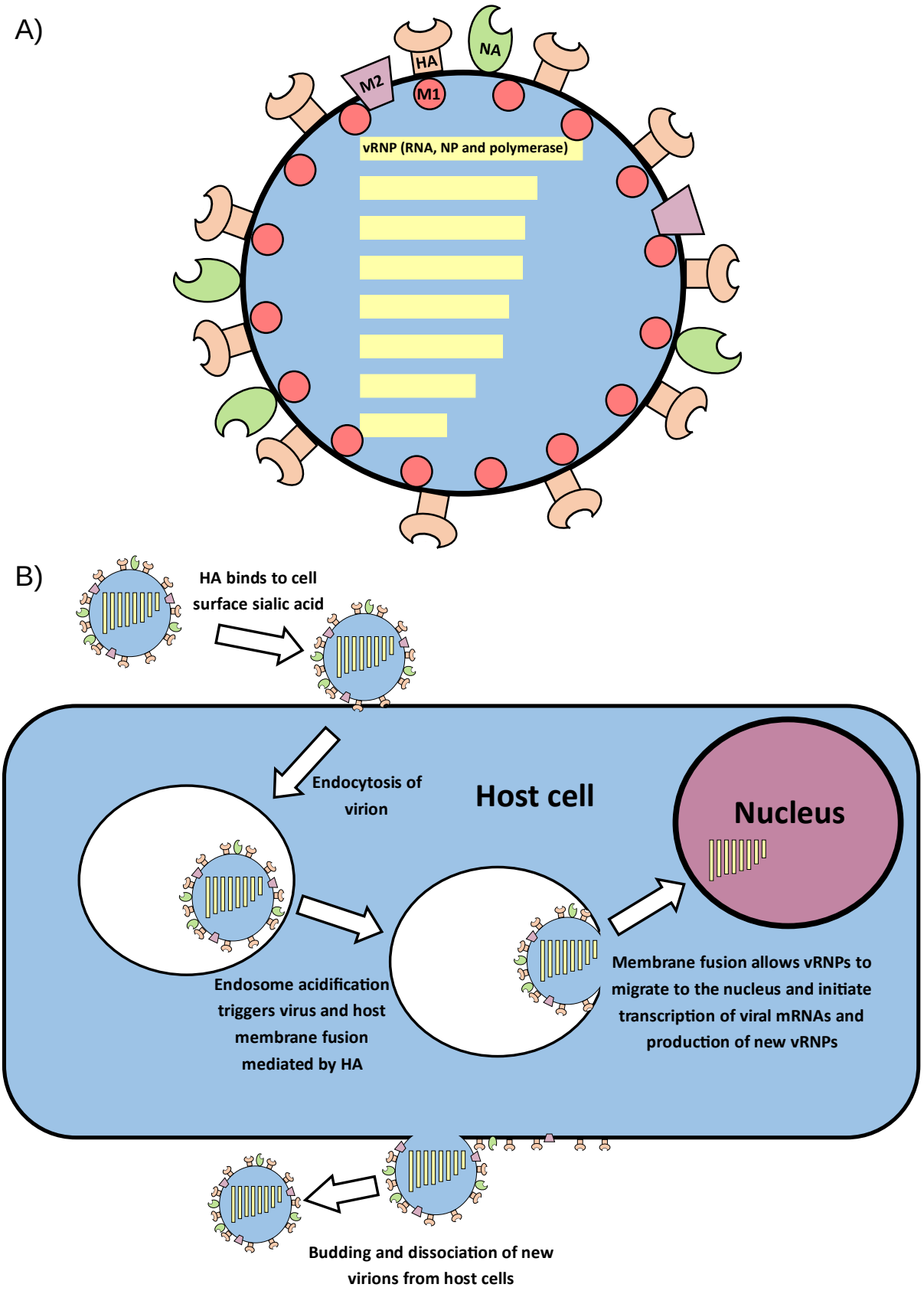


Figure INT 1a,b: Influenza A and B virion structure and life cycle A) Cartoon representation of the influenza virion. The membrane is populated with 3 viral proteins. HA (tan) mediates influenza virus entry into the host cell. NA (green) facilitates virion dissociation with the host cell after budding. M2 (purple) is a transmembrane proton channel involved in the virus uncoating process. M1 protein (red) associates with the inside surface of the virion membrane and interacts with the 8 viral ribonucleoproteins (vRNPs). vRNPs consist of a genome segment, influenza nucleoprotein (NP) and a copy of the heterotrimeric viral polymerase. **B)** Schematic of the influenza cellular life cycle outlining entry, uncoating and virion budding.

Influenza virions associate with the host cell surface by the binding of HA with sialic acid (SA) linked to galactose moieties found on a variety of membrane glycoproteins and glycolipids. A large number of HA-SA connections promotes the endocytosis of virions into host cells by the macropinocytosis and clathrin-mediated pathways before they are trafficked to the endosome[17]. Host cell endosome acidification triggers activation of the M2 transmembrane proton channels, acidifying the inside of the virion and causing dissociation of the vRNP complexes from M1[18]. Meanwhile, the drop in endosome pH also triggers conformational change in the HA trimers, drawing virion envelope and endosome membranes together, fusing the two membranes and releasing the vRNPs into the cell cytoplasm[19]. vRNPs are imported into the nucleus, aided by host factors importin α and importin β[16].

In a process primed by the cap of a host transcribed mRNA, the influenza polymerase produces viral mRNAs complementary to the negative sense genome segments, facilitating the synthesis of new influenza proteins including non-structural proteins 1 and 2 (NS1 and NS2) which are not found in the virion[20]. At a lower frequency, unprimed initiation of RNA polymerisation creates sense copies of genome segments with no mRNA cap. These segments are packaged almost identically to vRNPs and are referred to as cRNPs (copy ribonucleoproteins). In a nearly symmetrical process, cRNP complementary copies are made, forming new negative sense vRNPs[21].

New influenza virions are assembled at and bud from the plasma membrane. Envelope proteins HA, NA and M2 are translated into the endoplasmic reticulum and glycosylated in the Golgi apparatus before being trafficked to the external surface of the cell membrane. vRNPs are exported from the nucleus and trafficked to the cell surface aided by NS2, M1 and host factor Rab11[16].

Influenza virus budding occurs from plasma membrane “rafts” with altered lipid composition at which HA, NA and M2 concentrate. M1 binds to the tails of HA and NA on the internal surface of the membrane and vRNPs in turn associate with M1. The process of budding is still not fully understood and may occur by more than one redundant mechanism. It is hypothesised that HA and NA on the outer surface of the plasma membrane cause positive curvature while M1 creates negative curvature on the internal surface and M2 has a role in membrane scission[22, 23]. NA cleaves sialic acid linkages bound to HA, untethering virions from the infected host cell[15].

1.2. Haemagglutinin structure and function

The haemagglutinin trimer can be crudely divided into two regions. A membrane distal bulbous head region consisting of a mixture of beta sheets and alpha helices and a narrower membrane proximal stem region consisting mostly of alpha helices perpendicular to the membrane. HA1 extends into the HA stem region but the majority of the subunit is located in the head region. This subunit induces virion endocytosis through its SA binding site located on the membrane distal surface of the molecule. HA2 comprises most of the stem region and mediates viral envelope-endosome membrane fusion. HA2 contains the fusion peptide, an amphipathic sequence which, upon endosome acidification, is inserted into the endosome membrane to draw the two lipid bilayers together. The stem region also contains the HA1/HA2 cleavage site which facilitates the maturation of the HA trimer[15]. Figure INT 2a shows the structure of an HA0 trimer with these features annotated.

The pH mediated conformational transition of HA is highly dependent on the structure and chemical properties of the molecule. Acidification of the surrounding solution modulates the charge of several HA residues by protonation of amino acid side chains. As a result, electrostatic repulsive forces progressively widen the spacing between monomers in the head region and untether the fusion peptide from its neutral pH location. As the head region separates into 3 lobes, the centremost alpha helix of HA2 (helix B) extends outward in the membrane distal direction pushing the fusion peptide on its tip into the endosome

membrane[24, 25]. Once the fusion peptide is inserted in the endosome membrane, helix B folds downwards, bringing the fusion peptide towards the viral membrane anchor of HA and promoting the fusion of the two membranes[19]. This structural transition is illustrated in figure INT 2b.

This complex transition between two stable protein conformations tightly constrains the sequence of HA in regions important to membrane fusion, resulting in almost complete conservation of many residues within the stem region including the fusion peptide. Some central residues in the SA binding site are also highly conserved due to their essential function in the viral life cycle[26]. Figure INT 2c illustrates the regions of HA which are conserved and the regions which vary more between influenza isolates.

1.3. Phylogeny and pandemic threats from influenza A

Due to the frequent reassortment of genome segments between different influenza viruses, an absolute influenza phylogeny cannot be constructed[27]. However, the antigenicity of an influenza virus is determined overwhelmingly by the genome segments 4 and 6 which encode the HA and NA genes respectively. Figure INT 2d shows a phylogenetic tree of influenza A and influenza B HA. Influenza A HAs are classified into group 1 and group 2 and further partitioned into the subtypes H1-H18 while influenza B HAs belong to either the Victoria lineage or the Yamagata lineage. Similarly, the NA genes of influenza A are classified as subtypes N1-N11[28], so influenza A viruses are often referred to according to the subtypes of their primary antigenic determinants (e.g. H1N1).

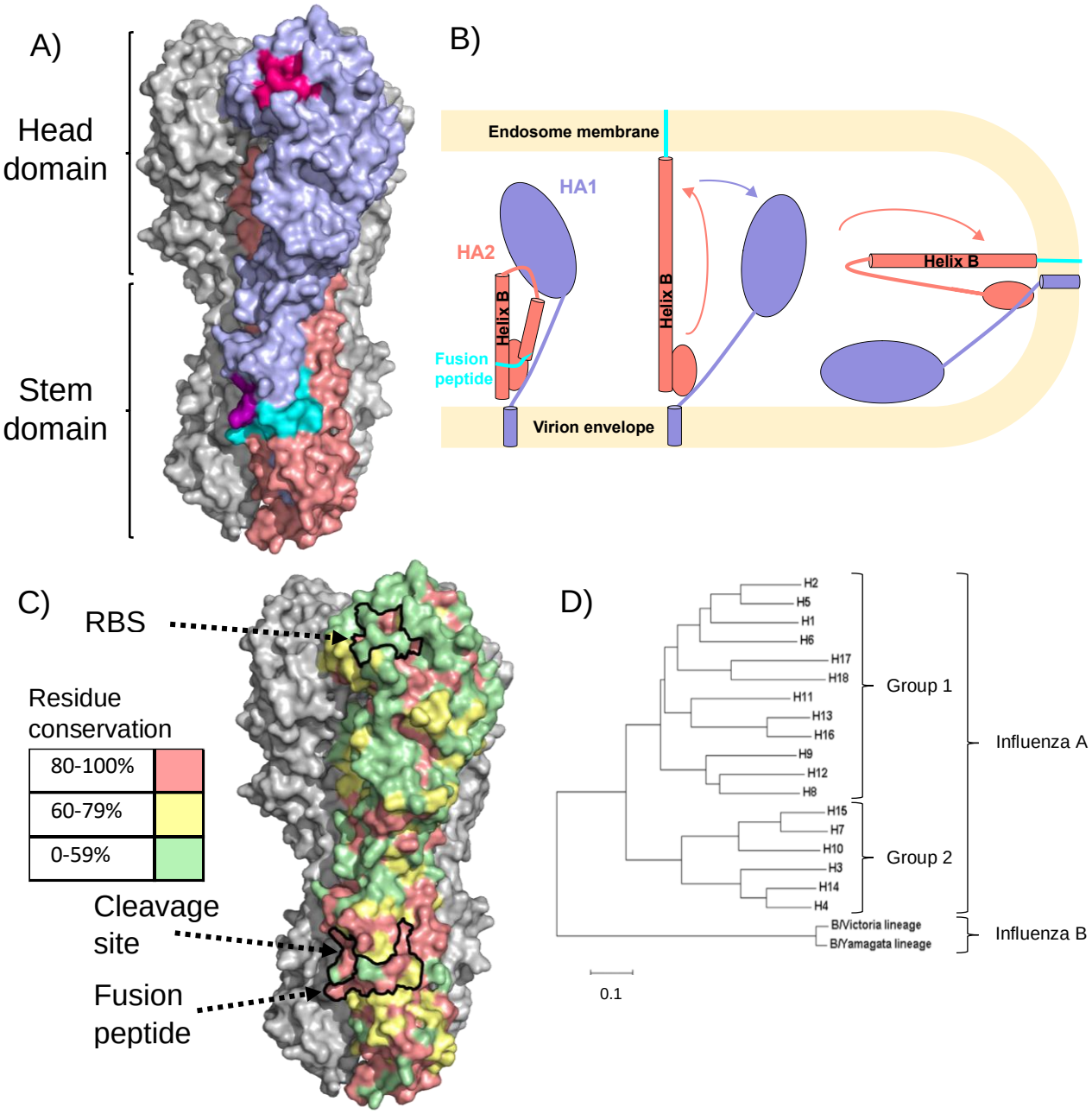


Figure INT 2a,b,c,d: Influenza haemagglutinin structure, conservation and phylogeny A) Structure of an H1 HA (A/California/04/2009, PDB code 3LZG)[29]. One unit of the trimer is highlighted, HA1 is coloured pale blue and HA2 is coloured pale red. The fusion peptide is highlighted in cyan. Upon pH mediated conformational change, it straightens and is inserted into the host membrane. The residues immediately surrounding the HA0 cleavage site are highlighted in purple. The receptor binding site is coloured pink. **B)** Schematic of the HA structural transition upon endosome acidification. Leftmost is the neutral pH conformation of the HA1 (blue) and HA2 (red) subunits together. Helix B is labelled and the fusion peptide is highlighted in cyan. The middle structure shows a low pH transition state with the fusion peptide inserted into the endosome membrane. The right structure shows the post fusion conformation with helix B inverted, bringing the two membranes together. **C)** The same HA structure as **A)** is used to display a heat map of influenza residue conservation across 4 influenza A group 1 pandemic risk subtypes (H1, H2, H5 and H9). 50 influenza HA sequences selected at random from each of these subtypes were downloaded from the influenza research database[30] (March 2020). Sequences were aligned and the frequency of the consensus residue at each position was calculated. Key residues in the RBS, fusion peptide and cleavage site are highly conserved as well as many residues in the stem region at the HA1/HA2 interface. Figure is modified from [31].

C) Phylogeny of HA of influenza A and B. Figure modified from [32]. The scale bar indicates amino acid substitutions per residue position.

The past 4 pandemic outbreaks of influenza A have been recent enough for the molecular characterisation and sequencing of the causative viruses. The first and most deadly of these pandemics was the 1918 “Spanish flu” caused by the spillover of an avian influenza H1N1 virus into the human population. An estimated 40-100 million people died in the outbreak and, unusually for influenza, many of the casualties were young and previously healthy[1]. In 1957, a reassortment event of an H1N1 virus descended from Spanish flu with another avian virus gave rise to a new pandemic. Both the HA and NA encoding genome segments of the previous H1N1 seasonal virus were replaced in the reassortment. The resultant reassorted virus was an H2N2 virus and was referred to as Asian flu. It was highly transmissible but much less pathogenic than the 1918 influenza strain, killing approximately 1 million people in the pandemic outbreak. Around a decade later, in 1968, a descendant of the H2N2 pandemic virus reassorted with an avian influenza virus to produce the H3N2 “Hong Kong” influenza pandemic virus. The most recent pandemic resulted from a complex series of reassortment events of avian, swine and human (Hong Kong 1968 descended) viruses. The reassortments occurred in livestock pigs and the resultant virus was transmitted to humans in 2009. It was an H1N1 virus which was referred to as “swine flu” or “(H1N1)pdm09” [33-35].

Current circulating seasonal influenza strains are H3N2 descendants of the 1968 pandemic strain and H1N1 descendants of the 2009 pandemic strain as well as representatives of the Victoria and Yamagata lineages of influenza B[36, 37].

Despite much research in this field, the exact determinants of transmissibility, virulence and pandemic risk in influenza strains remains largely unknown, so the genotype of the next influenza pandemic is hard to predict[34, 38]. Viruses with HA subtypes H1, H2 and H3 have a proven track record of causing pandemics so can be assumed to pose a risk for the future[34, 39]. In particular, the lack of H2 virus circulation in the human population since the 1960s has rendered the population under 50 years old more vulnerable to H2 viruses[40].

In addition, in recent years, predominantly in China, avian H5N1 and H7N9 viruses have infected humans with alarmingly high mortality, in the range 30-50%, although human to human transmission appears to be inefficient at present[33, 34, 41, 42]. Serological evidence suggests that H9N2 viruses are crossing over into human populations at greater frequency than H5 or H7 viruses but there is little evidence yet of high virulence from H9N2 viruses[43-45].

In summation, while it is prudent to prepare for potential influenza outbreaks from any of the antigenic subtypes of influenza, historical precedent and recent zoonotic spillovers suggest that the focus should be on pandemic preparedness for viruses of the HA subtypes H1, H2, H3, H5, H7 and H9[33]. Meanwhile, there is a continual need for improvement to the standard of prophylaxis and care for the four currently circulating seasonal influenza lineages (H1N1, H3N2, influenza B Victoria and influenza B Yamagata). One of the avenues of research aiming to lessen the impact of seasonal flu and prepare for future pandemic flu is the investigation of broadly neutralising antibodies[26, 46-48].

1.4. Antibodies to influenza

Currently, the most effective intervention against influenza outbreaks is an annual quadrivalent virus vaccine[49]. Months ahead of the winter flu season, a single strain from each of the 4 seasonal flu lineages is selected for vaccine inclusion, allowing time for virus growth, vaccine manufacture and testing[50, 51]. Recipients are primarily protected by a neutralising antibody response induced by the vaccine[52].

For several decades, it was believed that antibodies to influenza A virus would only neutralise a single strain or a narrow range of viral strains. Antibodies elicited by natural infection and vaccination are predominantly targeted to the highly diverse receptor-binding head region of HA[53-55]. Most of these antibodies neutralise influenza by directly blocking or sterically hindering the binding of HA to sialic acid[56]. Other mechanisms such as inhibition of viral egress by cross linking of viruses also contribute to antibody efficacy[57]. Antibodies inhibiting the HA-SA interaction can be detected with the haemagglutination inhibition (HAI) assay which was, until relatively recently, considered a gold standard method for

measurement of serum neutralising antibody titre and for differentiation between neutralising and non-neutralising influenza antibodies. An HAI positive antibody blocks the interaction of HA with SA and a negative result indicates that the interaction is unimpeded[58].

HA undergoes antigenic change every few years within human circulating strains, rendering the majority of antibodies elicited to previous vaccines and virus strains less effective or ineffective[5]. Furthermore, antigenic differences between subtypes are greater than those within subtypes. Greater still are the differences between group 1 antigens and group 2 antigens. Even in highly conserved regions, these two groups exhibit significant antigenic differences[28, 59].

These factors mean that vaccine-mediated protection is short term and lacking in breadth. Therefore, current vaccination strategies result in limited protection if strain selection predictions differ from the viruses circulating during the flu season. Pandemic influenza strains are, by definition, highly antigenically divergent from previously circulating stains so current vaccinations offer little to no protection[60-62]. In addition, the populations most vulnerable to influenza infection, such as the elderly, often struggle to produce high titre antibody responses to the vaccine, resulting in further reductions in vaccine effectiveness[63, 64].

1.4.1. Identification of cross-reactive antibodies

In 1993, using hybridoma screening technology, an antibody denoted C179 was identified which was able to neutralise both H1 and H2 subtype viruses[65]. In addition to a broader neutralisation range than previously characterised antibodies, C179 did not display HAI. The authors correctly hypothesised that the antibody C179 bound to the haemagglutinin stem region[66], a site on the protein with higher levels of inter-subtype conservation than the head region, explaining the cross reactivity[67]. Unfortunately, the implications of the discovery of C179 were overlooked in the field until several other stem binding broadly neutralising antibodies were reported in 2008[68, 69].

In the last decade and a half, broadly neutralising stem-binding antibodies have garnered significant scientific attention. Many of these antibodies have been discovered with reactivity profiles extending beyond group 1 subtypes to encompass group 2 subtypes and some are able to bind influenza B HA as well[47, 70-75].

Alternative target sites of cross-reactive influenza antibodies have also been identified. A small proportion of antibodies targeting the receptor binding site on haemagglutinin have been found to have subtype cross reactivity[76]. C05, for example, can neutralise a wide range of viruses including many from the subtypes H1, H2, H3, H9 and H12[77]. Like their strain-specific equivalents, broadly neutralising head-binding antibodies are HAI positive and neutralise influenza by blocking virus attachment to host cells[57]. More recently, an additional class of cross-reactive HA head antibody has been identified, targeting the conserved interface between trimer subunits. Though normally occluded, flexibility in the HA trimer can transiently reveal this epitope. These antibodies are non-neutralising but can be protective in vivo by promoting immune effector functions against infected cells[78-80].

The other influenza envelope proteins, NA and M2, can also be targeted by cross reactive antibodies, some with very broad reactivity profiles. For both of these influenza proteins, only a subset of the elicited cross-reactive antibodies have been shown to be neutralising, depending on the epitope targeted in each case. Neutralising anti-NA antibodies inhibit neuraminidase function and were therefore hypothesised to inhibit viral egress[81]. Likewise, neutralising anti-M2 antibodies were hypothesised to block the functionality of the M2 proton channel, disrupting viral core uncoating[82, 83].

1.4.2. Mechanisms of stem-binding antibody neutralisation

The area bound by broadly neutralising stem-binding antibodies consists of residues from both subunits of HA. It includes the fusion peptide as well as many of the residues involved in the conformational change which catalyses viral envelope-endosome membrane fusion[26]. Stem binding neutralising antibodies

stabilise the neutral pH pre-fusion conformation of the stem region under acidic conditions, inhibiting the conformational change upon endosome acidification.

Extensive evidence supports this theory. Complexes of the stem-binding antibody (CR6261) and the HA trimer formed at neutral pH maintain the neutral pH HA conformation when the buffer is acidified. This was shown by forming protein crystals of the antibody complex in acidified buffer and determining the structure using X-ray crystallography [67]. Furthermore, the first discovered stem binding antibody, C179, was also shown to inhibit HA mediated membrane fusion. Cells expressing HA on their plasma membranes form syncytia when pulsed with low pH culture medium due to the fusogenic action of HA. Pre-treatment with C179 inhibited syncytia formation[65]. The same was later demonstrated with 3 other stem-binders (F10, D8, A66)[69]. Finally, CR6261 and CR8020 were observed using fluorescent single particle tracking to co-localise with the virus throughout endocytosis and endosome acidification. Viral genes were not expressed in these cells suggesting that by remaining bound to HA, the antibody prevented productive infection of the cell[57].

In addition, a few stem-binding antibodies show evidence of another mechanism of neutralisation. Upon budding, the cleavage of immature HA0 into HA1 and HA2 is carried out by various proteases found in the virus host such as human airway trypsin-like protease and transmembrane protease serine S1 members 2 and 4 (TMPRSS-2 and 4). The antibodies F16v3 and CR8043 have been shown to inhibit trypsin-mediated cleavage of HA0[70, 74] and CR8020 can inhibit cleavage by trypsin and TMPRSS-2 and 4. On the other hand, CR6261 showed no inhibition of proteolytic cleavage. This additional mechanism increases the potency of these antibodies if virus is incubated with them pre-cleavage compared to post-cleavage[57].

Immune effector function activation appears to be another key mechanism by which HA stem binding antibodies function *in vivo* and has also been demonstrated *in vitro*. A broadly neutralising stem-binding antibody, 6F12, demonstrated Fc-dependent activation of NK cells, indicating a capacity to induce antibody-dependent cellular cytotoxicity (ADCC). Furthermore, at low antibody doses, a panel of 5 broadly neutralising stem binders (6F12, F16, 2G02, 2B06

and 1F02) with effector function inducing Fc-regions show enhanced protection of mice from influenza infection compared to those without effector functionality. The authors inferred that, at sub-saturating concentrations, the protective effect of stem binding monoclonal antibodies can be enhanced by immune system effector functions including ADCC *in vivo*[84].

However, there is some concern that, in some circumstances, stem-binding antibodies interacting with Fc-receptors on immune cells may exacerbate the pathology of an influenza infection. Non-neutralising antibodies elicited to a mismatched dengue virus serotype can promote the infection of macrophages and dendritic cells when antibody Fc-regions bind to Fc-receptors, inducing phagocytosis of the infectious virus[85, 86]. This phenomenon is referred to as antibody-dependent enhancement (ADE). ADE occurring with influenza antibodies cannot be excluded but substantial evidence of this phenomenon has not yet been observed[87, 88]. Cases of influenza antibodies increasing pathology in animal models have been attributable to antibodies which promote rather than inhibit endosome fusion and these mechanisms are Fc-region independent[89, 90].

1.5. Application of broadly neutralising antibodies to influenza treatment and vaccination

Given the narrow immunity induced by current vaccines, broadly neutralising monoclonal antibody-based drugs could improve the standard of influenza care[91]. These drugs could provide therapeutic and prophylactic protection against a wide range of strains of influenza for vulnerable populations.

Several stem binding antibodies are currently in clinical trials including MEDI8852, CR9114 and MHAA4549A. All 3 are human derived and influenza A group 1 and 2 cross neutralising. Clinical trials of stem binding monoclonal antibodies thus far have focused on therapeutic treatment of acute disease as long-term maintenance of a prophylactic dose would be prohibitively costly. Many of the experiments are trialling combination therapies of stem binding antibodies with additional drugs targeting other stages of the influenza life cycle. For example, both MEDI8852 and MHAA4549A are being administered in combination with oseltamivir (a small molecule NA inhibitor) and CR9114 is being

combined with a pair of anti-HA receptor binding site antibodies and favipiravir (a nucleoside analogue for viral polymerase inhibition). It is hoped that combination therapies will act synergistically to inhibit influenza infection more than any single treatment and the selection of drug resistant virus variants will be hampered by cotreatment with the other drugs in the cocktail working by orthogonal mechanisms. Some of these trials have reached as far as phase II so it is probable that monoclonal antibody treatments for influenza will begin to be approved in the next few years[92-94].

Knowledge of broadly neutralising epitopes of influenza virus proteins and the antibodies that bind to them is also being applied to development of more broadly protective influenza vaccines. These vaccines aim to elicit humoral responses resembling characterised monoclonal antibodies and often focus on the normally immunologically subdominant HA stem. For example, chimeric HA vaccination schedules involve repeated immunisation with antigens with divergent head regions but identical stem regions, aiming to disproportionately amplify and affinity mature stem binding antibodies[78]. Hyperglycosylated antigens aim to occlude highly variable epitopes to favour broadly cross reactive antibody responses to exposed conserved epitopes[79]. Another HA stem immunogen for a proposed vaccine is modified to bind more tightly to the germ line sequences of known broadly neutralising antibodies to elicit similar antibodies in patients[95]. These approaches show considerable promise but will require optimisation and lengthy clinical trials before they are available to the public[96-98].

1.6. Single domain antibodies

As more stem binding anti-influenza antibodies were isolated, a trend emerged that many of these antibodies used only the heavy chain domain to bind to HA or their binding mode was heavy chain dominant[47, 68, 69, 72]. Interestingly, this trend also applied to broadly neutralising antibodies to the HA receptor binding site[77, 99] and antibodies to HIV conserved epitopes[100, 101]. These trends suggested that, in many cases, light chain contribution is unnecessary for broad and potent neutralisation of antigenically variable pathogens. It has been hypothesised that the reason for this is that heavy chain only binding modes result in a smaller

epitope or fewer essential epitope contacts which is beneficial for binding to a wide range of variants of the same antigenic site[31, 69, 102, 103]. However, the methods used to isolate many of these monoclonal antibodies do not preserve natural heavy and light chain pairings. An alternative hypothesis is that heavy chains which require little or no help from the light chain to bind to antigen are more likely to be functional with a mismatched light chain and, therefore, are preferentially selected in protocols which do not preserve chain pairings[104].

Among mammals, camelids (camels, llamas and alpacas) are unique in that they naturally produce functional antibodies without a paired light chain, meaning that the binding unit of the antibody is a single domain (referred to as a VHH domain) of approximately 15kDa[105, 106]. These heavy chain only antibodies are produced and affinity matured alongside antibodies with a conventional domain structure in the camelid immune system[107] (see figure INT 3a,b). The VHH domain can be expressed as a stable protein with binding functionality intact, independent of the rest of the heavy chain (see figure INT 3c). In this form they are often referred to as single domain antibodies.

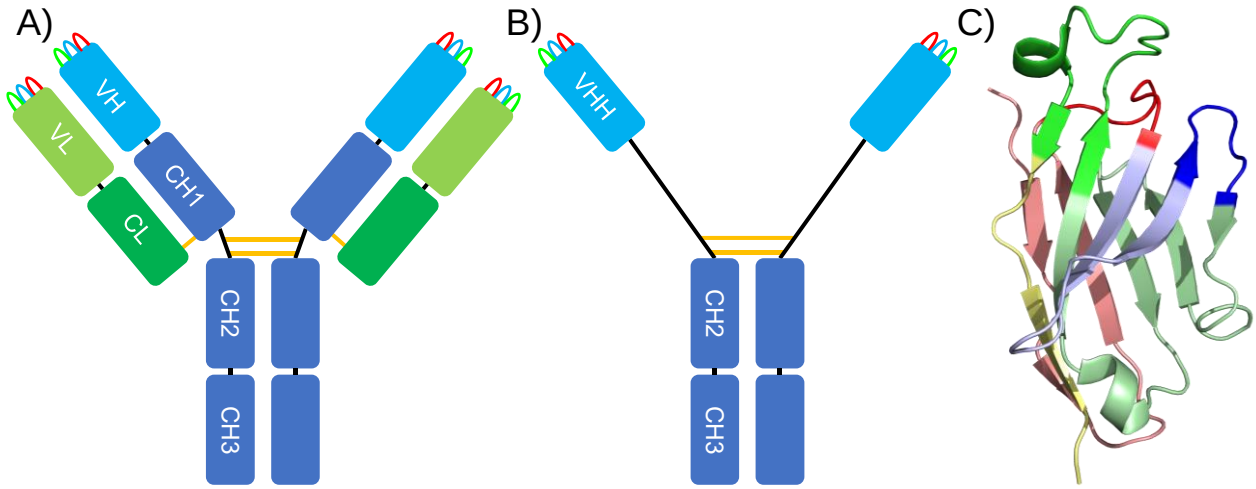


Figure INT 3a,b,c: Domain structure comparison of conventional and camelid heavy chain only antibodies A) Camelid IgG1 antibodies and all IgG antibodies of other mammals are composed of 4 peptides, two identical heavy chains (blue) and two identical light chains (green). The variable region determines the specificity of the antibody and consists of the heavy chain VH domain (light blue) and light chain VL domain (light green). Inter chain disulphide bonds link the 4 peptides and are depicted as orange lines. The constant region of the antibody comprises the CL, CH1, CH2 and CH3 domains. A subset of the constant region, the Fc region (CH2 and CH3 domains), interacts with the effector arm of the immune system, facilitating immune effector functions such as ADCC. **B)** IgG2 and IgG3 antibodies in camelids lack light chains and CH1 domains. The heavy chain variable domain (VHH domain) is therefore the only antigen-binding domain. Camelid IgG2 and IgG3 antibodies can elicit immune effector function due to a complete Fc region[107]. **C)** Cartoon structure of a VHH domain. Antigen contacting complementarity determining regions 1-3 (CDRs 1-3) are coloured red, blue and green respectively. Framework

regions 1-4 (FRs 1-4 - coloured pale red, pale blue, pale green and yellow) have less antigen contact and function to support and position the CDRs. The VHH shown is SD38 and the structure was solved from X-Ray crystallography data (PDB code 6CK8[108]).

VH domains of conventional antibodies contact the light chain VL domains largely with their FR2 region. This interaction is facilitated by hydrophobic FR2 residues on the outward facing domain surface at positions 37, 44, 45 and 47 (Kabat numbering system, which is used throughout this thesis[109]). VHH domains, lacking this VL interaction, encode a different set of amino acids at these residues, typically F37, E44, R45 and G47. These are referred to as hallmark residues and facilitate solubility and stability in solution[107, 110]. VHH domains, on average, have longer CDR3s than their VH counterparts. Longer CDR3s in VHH are often conformationally stabilised with an additional internal disulphide bridge between CDR1 and CDR3[111].

Characterisation of single domain antibodies has revealed that they can bind to a range of epitopes just as broad as conventional antibodies and with equivalent affinities[112]. The lack of a light chain paratope comes at a cost to binding surface diversity. However, analysis of VHH/antigen complex structures indicates that this lost paratope surface is compensated for by an increase in sequence and conformational diversity of the three CDRs[113] and increased use of the framework regions in the single domain paratope[114]. Because of their single binding domain and longer CDR3s, epitopes of VHHs are, on average, more compact and concave than those of conventional antibodies[115]. This biases VHH antibody responses towards enzyme active sites and other protein clefts[116, 117] and pathogen epitopes cryptic to conventional antibody responses[118, 119]. In addition, VHH domains show increased thermal and chemical stability compared to paired heavy and light chain constructs[107]. Unlike conventional antibodies, VHH domains express well with full functionality in bacteria, meaning production can be more cost effective[120, 121].

With the heavy chain bias of cross-reactive antiviral antibodies in mind, several studies have elicited and characterised camelid heavy chain only antibodies to influenza viruses. Camelid antibodies have been isolated to the HA stem[103, 108, 122], HA head region[103, 108, 122-125] and the other influenza envelope proteins[126-128].

1.6.1. Single domain antibody engineering

As lightweight, stable and modular binding units, VHH are ideally suited to multimerisation or fusion to other functional domains. Such VHH fusions have found application in research, diagnostics and as therapeutic candidates. For example, fusion of VHH to a fluorophore allows intracellular tracking of target proteins in real time. In one study, VHH fluorophore fusions binding to actin and proliferating cell nuclear antigen were expressed by zebrafish cells allowing precise monitoring of the cytoskeleton and cell proliferation[129]. In another application, fusion of a VHH to a bacterial homodimeric biotin-binding domain allowed easy immobilisation of the construct to a biotin coated surface. The surfaces with bound VHH were then used for surface plasmon resonance-based immunoassays[130]. As a proof of concept for an application of VHH antibodies to influenza prophylaxis, 4 influenza neutralising VHH domains were fused together with an antibody CH2 and CH3 domain. These fusion peptides dimerised to form homodimeric tetraspecific antibodies with an Fc-region capable of inducing immune effector functions. The small number of codons required to encode each VHH domain meant that the gene for this tetraspecific antibody could be encoded in a viral vector with very limited genome space[108].

In addition to the domain structure, the sequence of the VHH domain itself can be engineered for its intended application. Engineering of the VHH domain usually aims to improve the binding kinetics of the antibody[131-133] or maintain VHH binding kinetics under additional constraints such as high thermostability[134] or low immunogenicity[135]. Optimisation for low immunogenicity is desirable for non-human antibodies in therapeutic applications. Modification of the sequence and structure of a non-human antibody to resemble a human antibody (referred to as humanisation) reduces the risk of monoclonal antibody deactivation or adverse side effects mediated by the patient's immune system[136, 137].

The binding kinetics of an antibody and antigen in solution can be defined with two parameters. The on-rate (k_{on}) describes the rate of formation of antibody-antigen complexes and the off-rate (k_{off}) describes the rate at which complexes dissociate. Affinity (KD) is defined as k_{off} divided by k_{on} and a high affinity refers to a low value of KD. Generally, when engineering an antibody for improved

affinity, improvement in off-rate is far greater proportionally than the increase in on-rate achieved[132, 138-140]. Techniques used for antibody kinetics improvement are discussed in the first and second results chapters (sections 3 and 4).

1.7. R1a-B6

This project is focussed on a broadly neutralising stem-binding alpaca VHH, R1a-B6. R1a-B6 is cross reactive to influenza A group 1 HA antigens. Viruses of the subtypes H1 and H5 are neutralised with high potency by R1a-B6, H9 viruses are neutralised with much lower potency and low affinity binding to H2 antigens can be detected[102, 103, 141].

R1a-B6 was part of a panel of HAI negative antibodies affinity matured in and isolated from an alpaca. The alpaca immune system was stimulated with recombinant H1N1 2009 pandemic HA antigen (A/California/7/2009, produced using the baculovirus expression platform) with an adjuvant. B-cells were extracted from alpaca blood samples after an immune response had developed. Single domain antibody genes from B-cells were selected for H1 and H5 binding using a bacteriophage (phage) display library and from the selected libraries, genes were picked for characterisation. R1a-B6 was one of the broadest neutralisers of the VHH panel. Also identified in this panel was a clonal relative of R1a-B6 denoted R2b-E8 which cross neutralised H1 and H5 like R1a-B6 but had no detectable neutralisation of H9 viruses nor did it show detectable binding to H2 antigen[103].

Application of R1a-B6 has been explored in recent papers[141, 142]. In one application R1a-B6 was used in a proof of concept prophylactic gene therapy against influenza challenge. A gene encoding an R1a-B6 – mouse IgG Fc region fusion was packaged in an adeno-associated virus capsid. The viral vector was injected intramuscularly to infect mouse muscle cells and deposit its genetic material as a nuclear episome. The infected cells then stably expressed and secreted the bivalent antibody constructs into the bloodstream, protecting the mice from H1 and H5 influenza virus challenge. This construct was tested with an Fc region capable of inducing immune effector functions such as ADCC as well as an Fc incapable of such functionality. R1a-B6, as an IgG Fc dimer, was

sufficiently potent to be fully protective with and without immune effector function. This study demonstrates the potential for R1a-B6 to provide broad and long-term protection from influenza infection. The efficacy of this antibody without immune effector function may sidestep safety concerns about ADE and makes this technology more attractive as a vaccination alternative for the immunocompromised[87, 141].

In the other paper, R1a-B6 was used in an ELISA (enzyme linked immunosorbent assay) based vaccine potency assay capable of quantifying structurally native H1 or H5 HA in a sample relative to a reference standard. ELISA-based potency assays with broadly reactive antibodies require only hours to run and have the potential to replace the current gold standard seasonal influenza vaccine potency assay, single radial diffusion. Single radial diffusion requires sheep antiserum generated to the virus strains selected for the vaccine. Generation of antisera takes several weeks. Implementation of an assay which does not require strain specific antisera would allow vaccine strain selection to occur closer to the time of vaccine release. This may reduce the risk of a mismatch between vaccine influenza strains and influenza strains circulating during flu season[142].

The differences between the binding modes of conventional and single domain antibodies invites a comparison between HA stem binding antibodies of both types. While many structures exist for conventional antibodies in complex with HA, only 3 VHH-HA stem complexes have been solved[108]. In addition, epitope mapping of R1a-B6 on HA has identified substitutions at several HA stem residues capable of abrogating the binding interaction[102]. Figure INT 4a,b,c,d compares the epitope footprints of VHH and representative conventional antibodies with similar reactivity ranges. The number of sampled VHH is too low to draw any significant conclusions but it does highlight that footprints of many conventional and single domain antibodies, including R1a-B6, overlap substantially. This is unsurprising given the high level of functional conservation in that site on the HA.

Experiments characterising the interactions of these HAI negative VHH with the HA stem have demonstrated that pre-incubation of HA with VHH inhibits the acid mediated conformational change required for fusion. In addition, neutralising

stem binding VHH do not bind to the post fusion HA conformation. These results provide strong evidence that, like conventional stem binding antibodies, stem binding VHH neutralise influenza viruses by endosome membrane fusion inhibition[103, 108]. The HA0 protease inhibition mechanism observed in some conventional stem binding antibodies has not been reported on in the literature for single domain antibodies[57, 74].

Like their conventional counterparts, the *in vivo* efficacy of some stem binding VHH constructs is enhanced by immune effector function[108]. For R1a-B6, as mentioned above, there is no evidence for or against this phenomenon. The concentration of the bivalent R1a-B6-IgG Fc fusion present was too high to detect a differential effect between constructs with and without effector functionality[141].

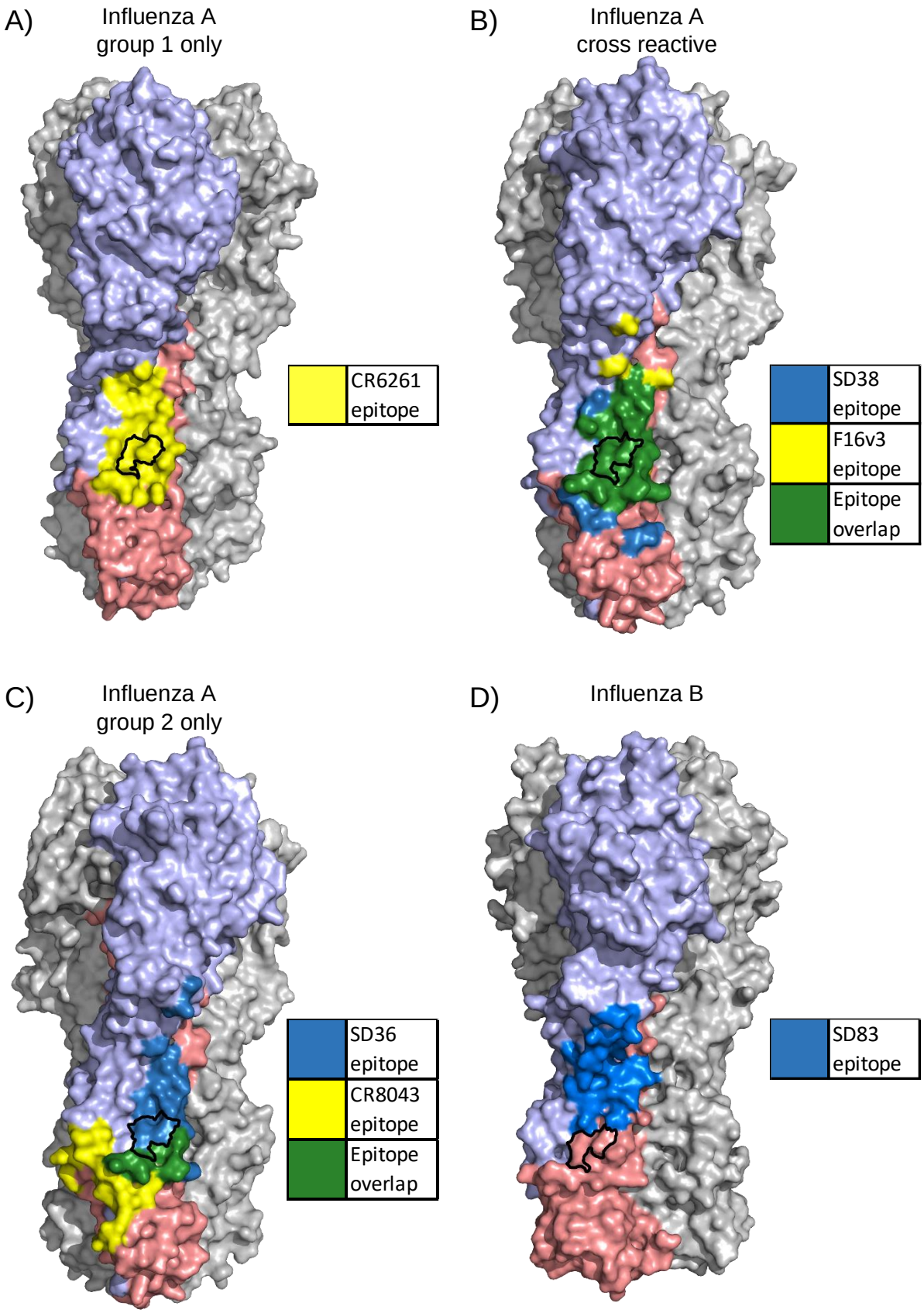


Figure INT 4a,b,c,d: Comparison of epitopes of stem binding conventional antibodies and single domain antibodies with similar neutralisation ranges Haemagglutinin trimer structures are used to display epitope footprints of single domain antibodies (blue), conventional antibodies (yellow) and areas of epitope overlap (green). Residues HA1 G20, W21 and I45 (H3 residue numbering) which are predicted to be inside the R1a-B6 epitope footprint are outlined in black[102].

A) Epitope of influenza A group 1 binding antibody CR6261 displayed on H1 (A/California/04/2009). **B)** SD38 and F16v3 (both pan influenza A neutralising) epitopes displayed on H3 (A/Hong Kong/1/1968 PDB code 4FNK)[77]. **C)** SD36 and CR8043 (both group 2 cross-neutralising) epitopes displayed on A/Hong Kong/1/1968 HA. SD38 and F16v3 have a high degree of overlap in their epitope footprints while SD36 and Group 2 stem binding conventional antibodies have sharply contrasting epitope footprints. This may indicate a difference in binding preference between conventional antibodies and sdAbs to the group 2 stem. **D)** SD83 (influenza B neutralising) epitope displayed on B/Brisbane/60/2008 (PDB code 4FQM)[73]. SD83 binds further up the stem than SD36 and SD38 and, unlike the other 2 VHH footprints, does not include residues HA2 20, 21 or 45. The residues included in all of the epitope footprints were determined based on protein structures in complex (PDB codes 6FYU, 6FYT, 6FYW, 3ZTJ, 4NM8, 3GBN)[67, 70, 74, 108] using the PBDePISA online tool[143].

1.8. Project aims and experimental approach

Given the epitope overlap of R1a-B6 with several antibodies with a greater neutralisation breadth, it was hypothesised that optimised variants of R1a-B6 may be able to bind effectively to HA antigens from a wider range of subtypes. Thereby, these variants would expand on the range of viruses R1a-B6 can potently neutralise. Expansion of neutralisation breadth was observed to occur during the *in vivo* affinity maturation of the stem binding antibody F16. Extrapolated ancestors of the F16 antibody could bind to group 1 HAs but not group 2 while the fully affinity matured antibody bound to all subtypes of influenza A HA tested[70]. Similar phenomena have been seen in which clonal lineages of anti-HIV and SARS-CoV2 antibodies broaden their neutralisation ranges over the course of *in vivo* affinity maturation[144, 145].

Little has been published on attempts to broaden the neutralisation range of antiviral antibodies *in vitro* and results of such attempts have been mixed. Randomisation of 6 residues in the heavy chain CDR3 and yeast display flow cytometric selection was used on influenza HA head binding antibody C05 to identify variants with strong binding to H1, H3 and H5 HA. However, it was found that substitutions improving H1 affinity reduced affinity for H3 HA and vice versa[146]. In a more recent study, a SARS-CoV1 neutralising VHH with weak cross reactivity to SARS-CoV2 was modified with amino acid substitutions suggested by analysis of an *in silico* model of the interaction. A combination of 3 substitutions was found which dramatically improved affinity of the VHH to SARS-

CoV2 viral antigen whilst maintaining the already high affinity for SARS-CoV1 viral antigen[138].

In this project we aimed to identify and characterise variants of R1a-B6 with a broader and more potent neutralisation profile than the original antibody using *in vitro* methods.

In the first results chapter (section 3), we gathered data on the interaction of R1a-B6 with HA. We used an alanine scan to infer which CDR residues on R1a-B6 were essential to the binding interaction and which could be substituted without loss of binding. An Illumina sequencing dataset was also mined for VHH sequences clonally related to R1a-B6. Data from both sources was used to design a combinatorial library of R1a-B6 variants to maximise the chances of finding functional improvement. This library was transformed into yeast for surface display and flow cytometric sorting for more broadly reactive variants.

Results chapter 2 (section 4) describes the flow cytometric sorting experiments used to enrich and isolate improved variants of R1a-B6 for further characterisation. Illumina sequencing was used to track the frequency of individual mutations and whole VHH clones through the rounds of selection.

In the next chapter (section 5) we expressed and purified R1a-B6 variants isolated from sorting experiments to characterise them in a variety of assays. R1a-B6 variants were compared to the unmodified antibody in surface plasmon resonance kinetics assays, pseudovirus neutralisation assays and ELISAs. HA antigens from a variety of subtypes were used to assess the breadth of reactivity of R1a-B6 and its variants.

The final results chapter (section 6) describes an epitope mapping experiment using a site saturated mutagenised HA library displayed on yeast. R1a-B6, two improved variants of R1a-B6 and R1a-B6 clonal relative R2b-E8 were used to stain the HA mutant library and unstained yeast cells were selected. Thereby, HA mutants with disrupted VHH binding were enriched. Illumina sequencing was applied to identify enriched point mutations and highlight the differences in epitope between the antibodies. This data was used to identify polymorphisms responsible for antigenic differences between HA of different subtypes.

2. Materials and methods

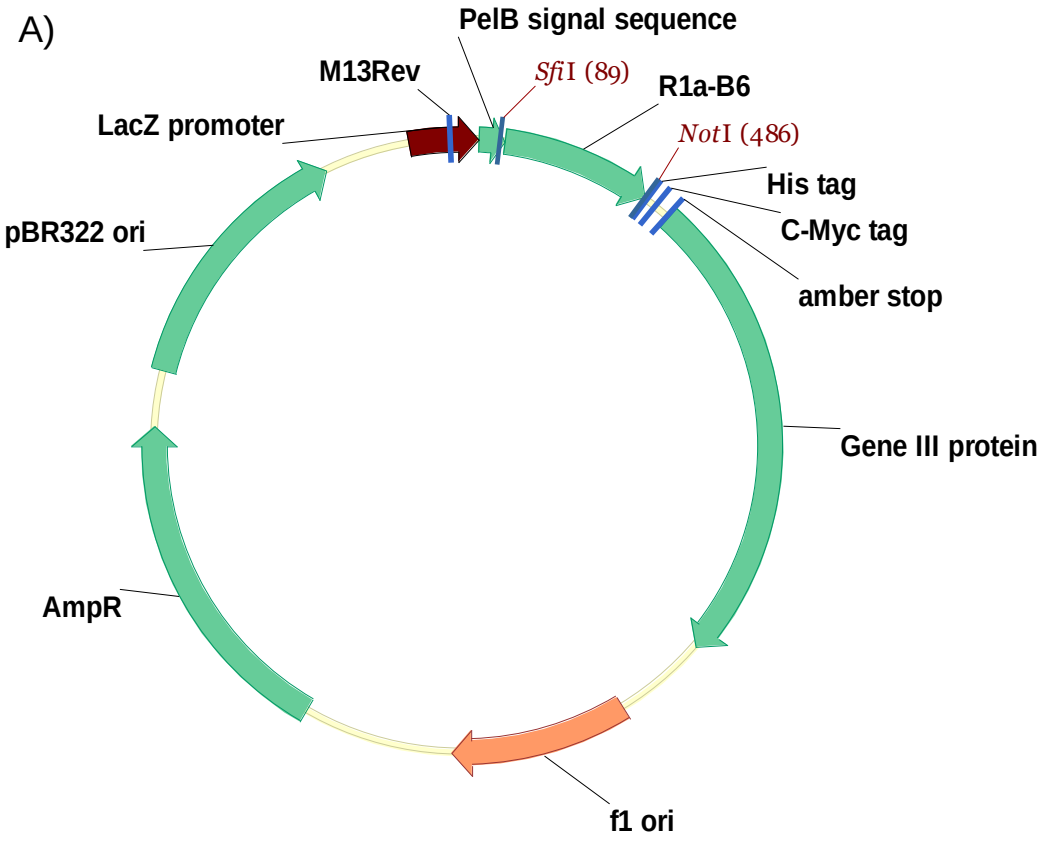
2.1. Molecular biology

2.1.1. pNIBS-1 and pNIBS-5 plasmids

Two plasmids were used in the experiments of this project: pNIBS-1[103] and pNIBS-5[102] (see figure MET 1a,b).

pNIBS-1 is a plasmid that can be used for phage display and soluble protein expression in the periplasm of *Escherichia coli* (*E.coli*) although it was used exclusively for soluble protein expression in this project.

pNIBS-5 is a single copy plasmid in EBY100 yeast[147] (a strain of *Saccharomyces cerevisiae*). pNIBS-5 facilitates inducible surface display of protein which can be used for yeast display assays and selections. pNIBS-5 can also be carried by *E.coli* where it confers beta-lactam resistance, allowing easy amplification and gene manipulation.



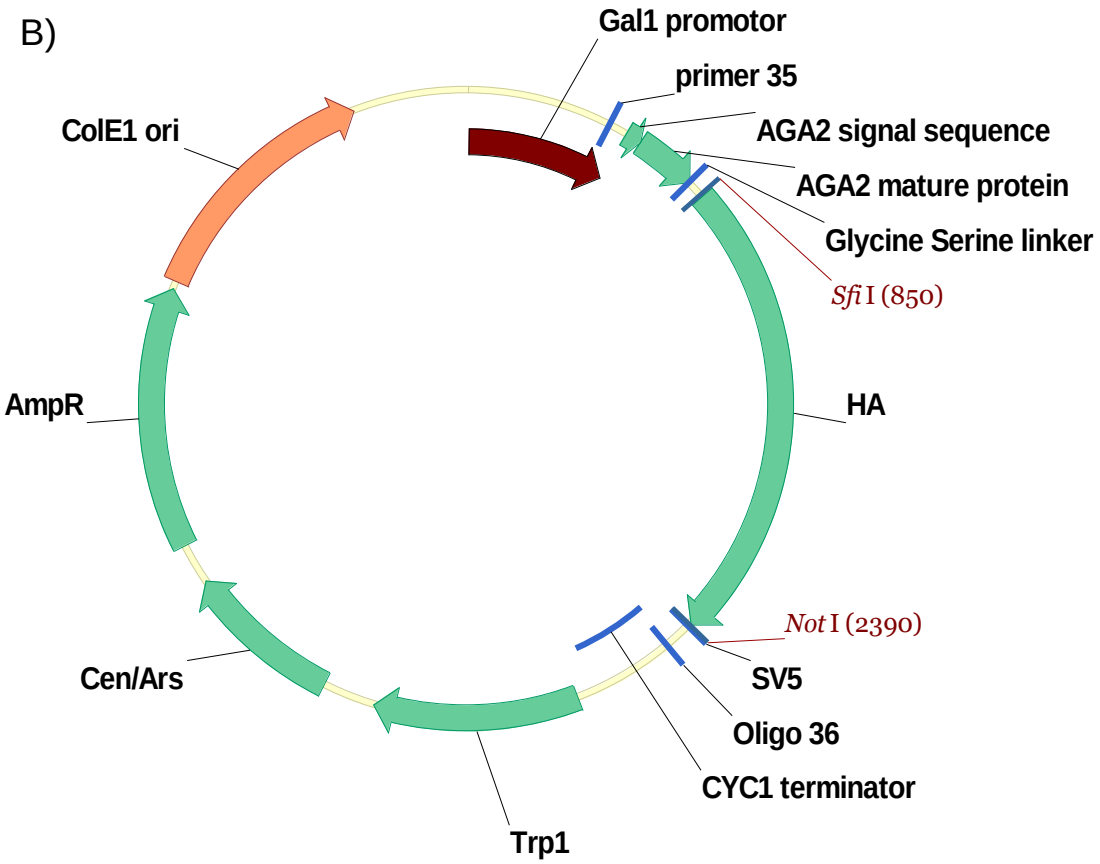


Figure MET 1a,b: Feature maps of plasmids used in this project A) pNIBS-1 plasmid with R1a-B6 gene – used for soluble VHH expression. As a result of the pBR322 ori sequence, pNIBS-1 is maintained at a copy number of around 20 in *E.coli* and the AmpR gene confers beta-lactam antibiotic resistance. When induced by isopropyl β-D-1-thiogalactopyranoside (IPTG) and free of dextrose repression, the LacZ promotor promotes the expression of a VHH domain (R1a-B6 in this case) with a PelB periplasmic signal sequence at the N-terminus and a His-tag followed by a Myc-tag at the C-terminus. In strains of *E.coli* with the phenotype glnV, such as TG1, an amber stop codon is sometimes read through, resulting in expression of VHH-Gene III fusions which can be used for phage display. The f1 ori allows the plasmid to be replicated and packaged into phage. The phage display functionality of this plasmid was not used in this project. WK6 and BL21 *E.coli* do not read through the amber stop resulting in only the shorter gene product being expressed. The annealing site of sequencing primer M13Rev is also annotated on pNIBS-1. **B)** pNIBS-5 plasmid with haemagglutinin gene – used for yeast display of VHH and HA. When carried by EBY100, the Cen/ARS (Centromere sequence/autonomously replicating sequence) element acts as an origin of replication and allows the plasmid to act as a small independent chromosome, maintaining 1 copy per cell. The TRP1 gene confers the ability to survive on tryptophan deficient media. The Gal 1 promoter, when induced by galactose and raffinose expresses the gene product: AGA2 signal sequence, AGA2 mature protein (which facilitates surface display by binding to AGA1), glycine-serine linker, surface expressed protein (HA in this case), SV5 tag. This product is displayed on the surface of the yeast cell. The *NotI* and *SfiI* sites allow the gene expressed on the yeast cell surface to be conveniently replaced. In *E.coli*, the ColE1 ori element is the origin of replication and causes pNIBS-5 to have copy number of approximately 20 per cell while the AmpR gene confers beta-lactam resistance. The annealing sites of the sequencing primers primer 35 and oligo 36 are annotated on to the plasmid too.

2.1.2. Growth, transformation and plasmid extraction of bacteria

E.coli liquid culture, unless stated otherwise, was grown up in 2xYT medium (16g/l bactotryptone (Gibco 211705), 5g/l NaCl, 10g/l yeast extract (Gibco 211929)) shaken at 225 rpm and 37°C. Small volumes of liquid culture (2-5ml) were grown in 14ml Falcon tubes (Corning 352059). Larger volumes were grown in various sizes of glass baffled flasks.

Solid cultures of *E.coli* were grown on 2xYT agar (2xYT medium plus 15g/l bacteriological agar (Sigma A5306)) petri dishes unless stated otherwise.

E.coli transformed with pNIBS-1 or pNIBS-5 plasmid were grown with 0.1µg/ml carbenicillin (Fisher BP2648) in both solid and liquid culture.

QIAprep Spin Miniprep Kits (Qiagen 27106) were used for bacterial plasmid extraction and purification according to manufacturer's instructions on 2-5ml of stationary phase culture. Plasmid was eluted into 30µl nuclease-free water (Invitrogen AM9938). When larger quantities of plasmid were needed, the QIAGEN Plasmid Midi Kit (Qiagen 12143) was used with 50ml of stationary phase culture according to manufacturer's instructions and the DNA pellet was dissolved in 200µl nuclease-free water.

Four strains of *E.coli* were used in this project: NEB5α (NEB C2987), BL21 (NEB C2530), WK6 (ATCC 47078) and XL1-Blue supercompetent cells (Agilent 200236). Plasmids were transformed into NEB5α, BL21 and XL1-Blue supercompetent cells by heat shock according to their respective manufacturers' protocols.

To prepare competent WK6 cells, a single colony of WK6 was inoculated from an agar plate into 50ml L-broth (10g/l bactotryptone, 10g/l NaCl, 5g/l yeast extract), shaken and incubated until the OD600 was 0.3-0.5 (see section 2.1.4). The culture was chilled on ice and then pelleted in a centrifuge (5000g, 10 minutes). The supernatant was discarded and the cells were resuspended in 5ml of ice cold sterile transformation solution (L-broth plus 10g/l $MgCl_2$, 100g/l

PEG4000 (Sigma 8.07490), 50ml/l dimethyl sulfoxide (DMSO) (Sigma D8418)). Competent WK6 cells were frozen at -80°C in 50µl aliquots.

To transform WK6 cells, plasmid DNA was added to a thawed aliquot of cells and incubated on ice for 30 minutes, followed by a 1 minute heat shock in a 42°C water bath and 3 minutes on ice. 250µl SOC medium (NEB B9020) was added and the mixture was shaken at 37°C for 1 hour before plating 100µl and 10µl on 2 2xYT plates with carbenicillin.

Purified plasmid was extracted from 3-5 transformed colonies and sequenced to find a colony that had been correctly transformed. Sanger sequencing was carried out by DBS Genomics (Department of Biosciences, Durham university, Stockton Rd, Durham, DH1 3LE, https://www.durham.ac.uk/departments/academic/biosciences/research/services/dbsgenomics/). For VHH genes on pNIBS-5 plasmid the primers Primer 35 and Oligo 36 were used. HA AmpnA Nest Fwd or HA AmpnB Nest Fwd were used for HA gene on pNIBS-5 and M13Rev was used for VHH genes on pNIBS-1 (table MET 1, Figure MET 2a,b,c).

Primer name	**Complementary to**	**Sequence**
Primer 35	pNIBS-5	GTAATACGACTCACTATAG
Oligo 36	pNIBS-5	GTCGATTTTGTTACATCTAC
M13Rev	pNIBS-1	CAGGAAACAGCTATGAC
HA AmpnA Nest Fwd	Gene encoding HA (A/California/7/2009)	GATCTGGTGGCGGAGGTTCTG
HA AmpnB Nest Fwd	Gene encoding HA (A/California/7/2009)	CAGTCCACGATTGCAATACAACTTG

Table MET 1: Primers used for Sanger sequencing

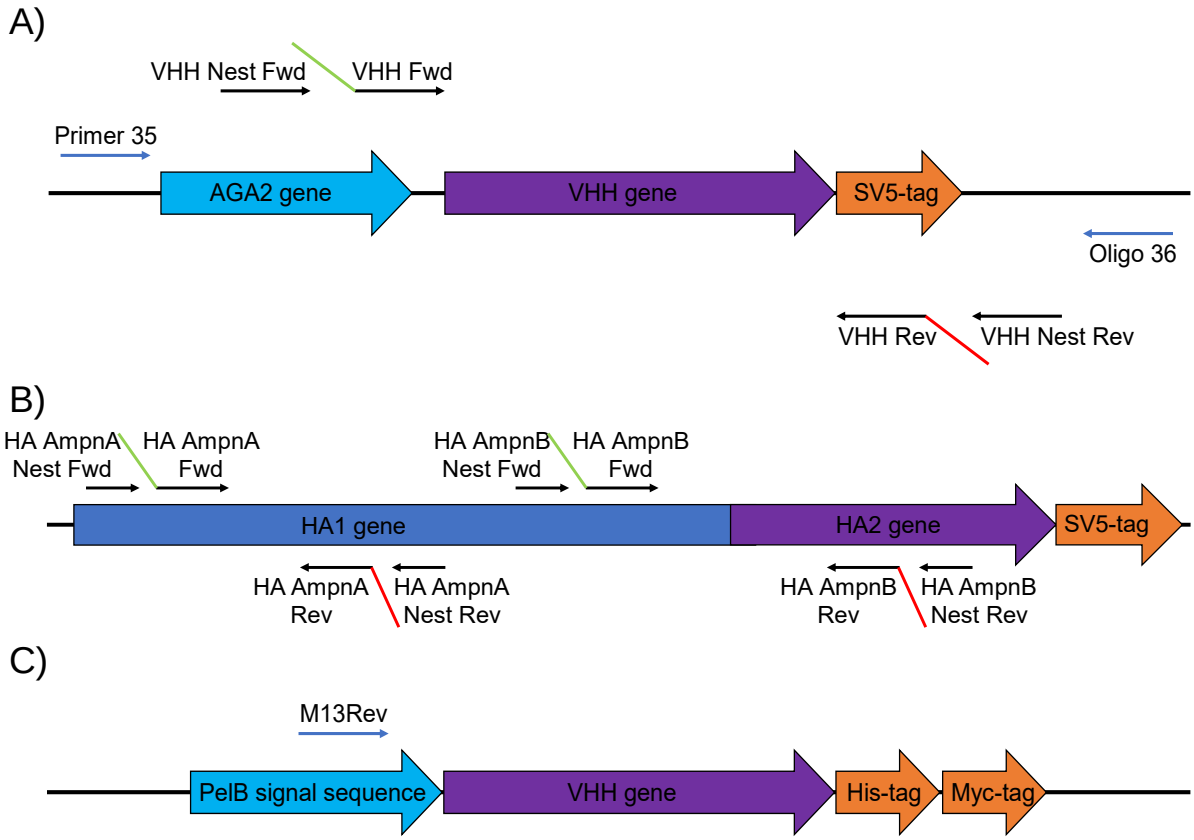


Figure MET 2a,b,c: Primer binding sites on plasmids pNIBS-1 and pNIBS-5 Primers binding to **A)** pNIBS-5 with a VHH gene insert **B)** pNIBS-5 with a HA gene insert **C)** pNIBS-1 with a VHH gene insert. Primers are shown in relative position to protein coding elements (not to scale). Sense primers are shown above the plasmid and antisense primers are shown below. Primers used for Illumina sequencing are shown as small black arrows. Illumina sequencing adapter sequences are shown in green (forward) and red (reverse). Other primers are shown as small blue arrows. Primer sequences are listed in table MET 1 and table MET 8a.

2.1.3. Yeast culture, plasmid extraction and transformation

Only the EBY100 strain of yeast was used in this project. EBY100 transformed with pNIBS-5 plasmid was grown in SD/CAA medium (5g/l casamino acids (BD 223050), 20g/l dextrose (Sigma G8270), 7g/l Yeast Nitrogen Base (Sigma Y0626), 10.19g/l (38mM) Na_2HPO_4-7H2O (Sigma S9390), 8.56g/l (62mM) NaH_2PO_4-H2O (Sigma S9638) filter sterilised, not autoclaved) shaken at 225 rpm and 30°C. Untransformed EBY100 was grown in YPD medium (50g/l of YPD broth powder (Sigma Y1375)). To prevent bacterial contamination, 0.1µg/ml of carbenicillin and 0.1µg/ml of kanamycin was added to liquid cultures. Small volumes of liquid culture (2-5ml) were grown in 14ml Falcon tubes. Larger volumes were grown in various sizes of glass baffled flasks.

EBY100 without plasmid was grown on YPD agar (50g/l of YPD broth powder, 20 g/l bacteriological agar) plates. SD/CAA agar (5 g/l casamino acid, 7 g/l Yeast Nitrogen Base, 15g/l bacteriological agar, 20 g/l dextrose (Sigma G8270) (33ml/l sterile 60% dextrose solution (w/v) added after autoclaving)) plates were used to grow yeast transformed with pNIBS-5. Both types of plates were incubated at 30°C.

Plasmids were extracted from 1.5ml of yeast liquid culture with the Zymoprep Yeast Plasmid Miniprep II kit (Zymo research D2004) according to manufacturer's instructions and eluted into 10µl nuclease-free water.

To transform yeast, the Sigma yeast transformation kit (Sigma YEAST1) was used. Unless stated otherwise, the short protocol was used. In several experiments, the ability of yeast to recombine homologous DNA was taken advantage of to clone antibody genes into the surface display open reading frame of pNIBS-5. 100ng pNIBS-5 backbone (the insert between the *Sfi*I and *Not*I restriction sites removed) and 100ng of DNA encoding the antibody were used in the kit protocol instead of the 1µg plasmid which was recommended. The inserts overlapped with the backbone with the sequences TGGTGGCGGAGGTTCTGCGGCCCAGC (synthetic double stranded DNA fragments) or GATCTGGTGGCGGAGGTTCTGCGGCCCAGC (R1a-B6 variant library amplified product) at the 5' end and GGCCGCTGGTAAGCCAATC CCAAACC (synthetic double stranded DNA fragments) or GGCCGCTGG TAAGCCAATCCCAAACCCATT (R1a-B6 variant library amplified product) at the 3' end. 100µl of the resuspended transformed yeast were plated on SD/CAA agar plates

Purified plasmid was extracted from 3-5 transformed colonies as described above. The resultant DNA from this extraction was too dilute for sequencing so NEB5α bacteria were transformed with the plasmid extract, grown to stationary phase and higher concentration purified plasmid was extracted and sequenced. Alternatively, in some instances, the insert region of pNIBS-5 was amplified by PCR using Primer 35 and Oligo 36 as primers (see section 2.1.10). Purified PCR product was sequenced. In both cases Primer 35 and Oligo 36 were used as primers for sequencing (see table MET 1 and figure MET 2a).

2.1.4. Measuring optical density of bacteria and yeast

Several protocols used in this project require measurement of OD600 of bacterial or yeast liquid cultures. This was done on an Eppendorf Biophotometer Plus in 2ml disposable cuvettes (Somatco 7591 15). Where density of yeast cells is estimated, this was done by OD600 measurement with the approximation that 2 x 10^7 yeast cells/ml have an OD600 of 1.

2.1.5. Storage of yeast and bacterial isolates and libraries

1ml of stationary phase culture was mixed with 0.5 ml filter sterilised 60% glycerol solution (v/v) in a 2ml Cryogenic Vial (Corning 430489) and frozen at -80°C.

2.1.6. Quantification of DNA and protein

DNA and protein were quantified on a NanoDrop 2000 (Thermo). When quantifying protein, the values from the NanoDrop were adjusted according to the A280 extinction coefficient calculated at “https://www.novoprolabs.com/tools /protein-extinction-coefficient-calculation”.

2.1.7. DNA gel electrophoresis and extraction of products run on gels

DNA was run in gels composed of 1% agarose (w/v) (Promega V3121) in TAE buffer (Thermo B49) and containing 1:10,000 GelRed Nucleic Acid Gel Stain (Biotium BT41003) Gels were run at 100V with 100bp (NEB N3231S) or 1000bp ladders (NEB N0552S) and samples loaded with 6x blue/orange loading dye (Promega G1881).

DNA in gels was visualised with a Syngene GelVue and gels were photographed through an orange filter.

When a band was to be cut out of the gel and purified, an Invitrogen Safe Imager 2.0 Blue-Light Transilluminator was used to visualise DNA and the band was cut out of the gel using a sterile scalpel blade. The DNA was extracted using a

QIAquick Gel Extraction Kit (Qiagen 28706) according to manufacturer's instructions and eluted from the column into 30µl nuclease-free water.

2.1.8. Restriction digestion

Restriction digestion reactions were carried out using the enzymes *Not*I HF (NEB R3189) and *Sfi*I (NEB R0123) at DNA concentrations from 0.2 to 3 µg per 50µl reaction volume. 5µl of Cutsmart buffer (NEB B6004) and 1µl/20 units (*Not*I HF) and 0.25µl/5 units (*Sfi*I) of enzyme at supplied concentration were used per 50µl reaction. For *Not*I *HF* digestions the reaction was incubated at 37°C for 30 minutes and *Sfi*I digestions were incubated at 50°C for 3 hours. Where both restriction enzymes were being used to digest the same DNA in the same reaction the 37°C step preceded the 50°C step.

Restriction reactions were purified using MinElute Reaction Cleanup Kits (Qiagen 28206) and eluted into 10µl nuclease-free water.

Inserts and backbones of pNIBS-1 and pNIBS-5 plasmids were extracted with a dual enzyme restriction reaction followed by running the product on an agarose gel then cutting out and purifying the appropriate band.

2.1.9. Ligation

To ligate a double digested insert and double digested plasmid backbone, the NEB quick ligation kit (NEB M2200) was used according to manufacturer instructions with a vector:insert molar ratio of 1:3. The completed reaction was kept on ice until 2.5µl of the 20µl reaction was used to transform competent *E.coli*.

2.1.10. Amplification by Polymerase chain reaction (PCR)

Unless stated otherwise, all PCR reactions were carried out to amplify VHH inserts on pNIBS-5 plasmid using the Primer 35 and Oligo 36 plasmids (see table MET 1 and figure MET 2a). The Phusion Flash High-Fidelity PCR Master Mix (Thermo F548) was used for these reactions. Reactions contained 0.5µM of forward and reverse primers and comprised 50% by volume Phusion master mix.

Template concentration varied. Total reaction volume was scaled according to the quantity of amplified DNA required. Thermal cycles recommended by the manufacturer were followed with an annealing temperature of 52°C and 30 cycles. DNA in completed PCR reactions was purified using a Qiaquick PCR purification kit (Qiagen 28104) according to manufacturer instructions. DNA was eluted into 30µl nuclease-free water.

2.1.11. Synthetic DNA

All synthetic DNA (single and double stranded) used in this project was ordered from Integrated DNA Technologies (IDT). DNA was ordered as a lyophilised powder and each oligonucleotide was resuspended in nuclease-free water to 100µM upon receipt. DNA was ordered with standard desalting (without additional purification steps) unless specified otherwise.

2.2. Manipulation of R1a-B6 and HA variants expressed on yeast

In several experiments, mutant R1a-B6 or HA variants were characterised by yeast display. In all cases, variants were encoded on the pNIBS-5 plasmid. The modification of wild type sequence and insertion of mutant genes into the pNIBS-5 plasmid is described in this section.

2.2.1. Site directed mutagenesis

Site directed mutagenesis of pNIBS-5 plasmid was achieved with the QuikChange II Site directed mutagenesis kit (Agilent 200523). Pairs of primers (table MET 2a,b) were designed to anneal to the sequence of either the R1a-B6 gene[103] or the HA (A/California/7/2009) gene around the site of mutation and to differ from the wild type sequence only at residues intended to be mutated. Mutagenesis reactions were performed on 50ng purified pNIBS-5 plasmids encoding the gene for either wild type R1a-B6 or wild type HA (A/California/7/2009). Primers were ordered with polyacrylamide gel electrophoresis (PAGE) purification according to the manufacturer's recommendation. Plasmids were mutagenised and amplified, then transformed

into the bacteria supplied with the kit (XL1-Blue supercompetent cells) according to manufacturer's protocol. Individual transformed bacterial colonies were picked (5 per mutagenesis reaction) and plasmid was extracted and sequenced. Purified plasmid extract containing the correctly mutagenised VHH or HA gene was used to transform EBY100 yeast. After experiments, plasmids in experimental yeast populations were extracted and sequenced to ensure data had been collected on the correct VHH and HA variants.

This method was used to insert alanine scan mutations into R1a-B6 and to insert mutations into HA based on epitope mapping results. Table MET 2a,b show the primers used in mutagenesis reactions.

A)

Primer name	**Mutation**	**Sequence (mutations in red)**
B6 CDR1 Gly26Fw	G26A	CTGCTGAAGAAGCTTGCAGAGGCTGCACAGG
B6 CDR1 Gly26Rv		CCTGTGCAGCCTCTGCAAGCTTCTTCAGCAG
B6 CDR1 Ser27Fw	S27A	GATATCTGCTGAAGAAGGCTCCAGAGGCTGCACAGG
B6 CDR1 Ser27Rv		CCTGTGCAGCCTCTGGAGCCTTCTTCAGCAGATATC
B6 CDR1 Phe28Fw	F28A	GCGATATCTGCTGAAGGCGCTTCCAGAGGCTGCAC
B6 CDR1 Phe28Rv		GTGCAGCCTCTGGAAGCGCCTTCAGCAGATATCGC
B6 CDR1 Phe29Fw	F29A	CCATGCGATATCTGCTGGCGAAGCTTCCAGAGGCTG
B6 CDR1 Phe29Rv		CAGCCTCTGGAAGCTTCGCCAGCAGATATCGCATGG
B6 CDR1 Ser30Fw	S30A	GCCCATGCGATATCTGGCGAAGAAGCTTCCAGAGG
B6 CDR1 Ser30Rv		CCTCTGGAAGCTTCTTCGCCAGATATCGCATGGGC
B6 CDR1 Arg31Fw	R31A	CCAGCCCATGCGATATGCGCTGAAGAAGCTTCCAG
B6 CDR1 Arg31Rv		CTGGAAGCTTCTTCAGCGCATATCGCATGGGCTGG
B6 CDR1 Tyr32Fw	Y32A	CAGCCCATGCGAGCTCTGCTGAAGAAGCTTCCAGAGG
B6 CDR1 Tyr32Rv		CCTCTGGAAGCTTCTTCAGCAGAGCTCGCATGGGCTG
B6 CDR1 Arg33Fw	R33A	GTACCAGCCCATGGCATATCTGCTGAAGAAGCTTCCAG
B6 CDR1 Arg33Rv		CTGGAAGCTTCTTCAGCAGATATGCCATGGGCTGGTAC
B6 CDR2 Ile51Fw	I51A	CTACCATCATAAGCGGCAGATGCGACCAACTCGCGCTG
B6 CDR2 Ile51Rv		CAGCGCGAGTTGGTCGCATCTGCCGCTTATGATGGTAG
B6 CDR2 Tyr53Fw	Y53A	GCTTGTACTACCATCAGCAGCGATAGATGCGACCAACTCG
B6 CDR2 Tyr53Rv		CGAGTTGGTCGCATCTATCGCTGCTGATGGTAGTACAAGC

B6 CDR2 Asp54Fw	D54A	GCATAGCTTGTACTACCAGCATAAGCGATAGATGCG
B6 CDR2 Asp54Rv		CGCATCTATCGCTTATGCTGGTAGTACAAGCTATGC
B6 CDR2 Gly55Fw	G55A	GTCTGCATAGCTTGTACTAGCATCATAAGCGATAGATGC
B6 CDR2 Gly55Rv		GCATCTATCGCTTATGATGCTAGTACAAGCTATGCAGAC
B6 CDR2 Ser56Fw	S56A	GGGGTCTGCATAGCTTGTAGCACCATCATAAGCGATAG
B6 CDR2 Ser56Rv		CTATCGCTTATGATGGTGCTACAAGCTATGCAGACCCC
B6 CDR2 Thr57Fw	T57A	GGGGTCTGCATAGCTTGCACTACCATCATAAGCG
B6 CDR2 Thr57Rv		CGCTTATGATGGTAGTGCAAGCTATGCAGACCCC
B6 CDR3 Asn93Fw	N93A	CCCCGGCGGATCTAAAGCACAGTAATAGACGGCCGTATC
B6 CDR3 Asn93Rv		GATACGGCCGTCTATTACTGTGCTTTAGATCCGCCGGGG
B6 CDR3 Leu94Fw	L94A	CCCCGGCGGATCTGCATTACAGTAATAGACGGCCGTATC
B6 CDR3 Leu94Rv		GATACGGCCGTCTATTACTGTAATGCAGATCCGCCGGGG
B6 CDR3 Asp95Fw	D95A	GAATCCCCGGCGGAGCTAAATTACAGTAATAGACGG
B6 CDR3 Asp95Rv		CCGTCTATTACTGTAATTTAGCTCCGCCGGGGATTC
B6 CDR3 Pro96Fw	P96A	GAATCCCCGGCGCATCTAAATTACAGTAATAGACGG
B6 CDR3 Pro96Rv		CCGTCTATTACTGTAATTTAGATGCGCCGGGGATTC
B6 CDR3 Pro97Fw	P97A	CCAGTATAGAATCCCCGCCGGATCTAAATTACAG
B6 CDR3 Pro97Rv		CTGTAATTTAGATCCGGCGGGGATTCTATACTGG
B6 CDR3 Gly98Fw	G98A	GCCCCAGTATAGAATCGCCGGCGGATCTAAATTAC
B6 CDR3 Gly98Rv		GTAATTTAGATCCGCCGGCGATTCTATACTGGGGC
B6 CDR3 Ile99Fw	I99A	CTGGCCCCAGTATAGAGCCCCCGGCGGATCTAAA
B6 CDR3 Ile99Rv		TTTAGATCCGCCGGGGGCTCTATACTGGGGCCAG
B6 CDR3 Leu101Fw	L101A	CCTGGCCCCAGTATGCAATCCCCGGCGGATC
B6 CDR3 Leu101Rv		GATCCGCCGGGGATTGCATACTGGGGCCAGG
B6 CDR3 Tyr102Fw	Y102A	CTGGCCCCAGGCTAGAATCCCCGGCGGATC
B6 CDR3 Tyr102Rv		GATCCGCCGGGGATTCTAGCCTGGGGCCAG

B)

Primer name	Mutation	Sequence (mutations in red)
HA1 H38Q Fwd	H38Q	CAGTACTAGAAAAGAATGTAACAGTAACACAGTCTGTTAACCTTTTAG
HA1 H38Q Rev		CTAAAAGGTTAACAGACTGTGTTACTGTTACATTCTTTTCTAGTACTG
HA1 V40K Fwd	V40K	GAAAAGAATGTAACAGTAACACACTCTAAGAACCTTTTAGAAGACAAGCATAACGGG
HA1 V40K Rev		CCCGTTATGCTTGTCTTCTAAAAGGTTCTTAGAGTGTGTTACTGTTACATTCTTTTC
HA1 V40Q Fwd	V40Q	GAAAAGAATGTAACAGTAACACACTCTCAGAACCTTTTAGAAGACAAGCATAACGGG
HA1 V40Q Rev		CCCGTTATGCTTGTCTTCTAAAAGGTTCTGAGAGTGTGTTACTGTTACATTCTTTTC
HA1 P299K Fwd	P299K	CAGCCTCCCATTTCAGAATATACATAAGATCACAATTGGAAAATGTCCAAAA
HA1 P299K Rev		TTTTGGACATTTTCCAATTGTGATCTTATGTATATTCTGAAATGGGAGGCTG

Table MET 2a,b: Primers used for site directed mutagenesis Forward and reverse primer pairs were designed to be almost exactly complementary to reverse and forwards strands of genes encoding **A)** R1a-B6 or **B)** HA (A/California/7/2009). Bases not complementary to their respective sequences are displayed in red. The changes in red are incorporated into the gene and result in the amino acid substitutions listed when the protein is expressed.

2.2.2. Whole gene synthesis and transformation into pNIBS-5

Other variants of R1a-B6 were ordered as synthetic double stranded DNA fragments (Sequences shown in table MET 3) and transformed into yeast with pNIBS-5 backbone using the homologous recombination method described in section 2.1.3.

Antibody variant name	Mutation compared to R1a-B6	Sequence (homology with pNIBS-5 plasmid highlighted and mutated codons in red)
VarB	FR3 73NANT->73NAKNT	TGGTGGCGGAGGTTCTGCGGCCCAGCCGGCCATGGCACAGGTGCAGCTCGTGGAGTCGGG CGGAGGCTTGGTGCAGCCTGGGGGGTCTCTGAGACTCTCCTGTGCAGCCTCTGGAAGCTT CTTCAGCAGATATCGCATGGGCTGGTACCGCCAGGCTCCAGGGGAGCAGCGCGAGTTGGT CGCATCTATCGCTTATGATGGTAGTACAAGCTATGCAGACCCCGTGAAGGGCCGATTCAC CATCTCCAGAGACAACGCCAAGAACACGGTGCATCTGCAAATGTACAGTCTGAAACCTGA CGATACGGCCGTCTATTACTGTAATTTAGATCCGCCGGGGATTCTATACTGGGGCCAGGG GACCCAGGTCACCGTCTCCTCAGCGGCCGCTGGTAAGCCAATCCCAAACC
VarC	FR3 73NANT->73NAENT	TGGTGGCGGAGGTTCTGCGGCCCAGCCGGCCATGGCACAGGTGCAGCTCGTGGAGTCGGG CGGAGGCTTGGTGCAGCCTGGGGGGTCTCTGAGACTCTCCTGTGCAGCCTCTGGAAGCTT CTTCAGCAGATATCGCATGGGCTGGTACCGCCAGGCTCCAGGGGAGCAGCGCGAGTTGGT CGCATCTATCGCTTATGATGGTAGTACAAGCTATGCAGACCCCGTGAAGGGCCGATTCAC CATCTCCAGAGACAACGCCGAGAACACGGTGCATCTGCAAATGTACAGTCTGAAACCTGA CGATACGGCCGTCTATTACTGTAATTTAGATCCGCCGGGGATTCTATACTGGGGCCAGGG GACCCAGGTCACCGTCTCCTCAGCGGCCGCTGGTAAGCCAATCCCAAACC
VarD	CDR3 93NLDPPGILY->93NLNPPGILY	TGGTGGCGGAGGTTCTGCGGCCCAGCCGGCCATGGCACAGGTGCAGCTCGTGGAGTCGGG CGGAGGCTTGGTGCAGCCTGGGGGGTCTCTGAGACTCTCCTGTGCAGCCTCTGGAAGCTT CTTCAGCAGATATCGCATGGGCTGGTACCGCCAGGCTCCAGGGGAGCAGCGCGAGTTGGT CGCATCTATCGCTTATGATGGTAGTACAAGCTATGCAGACCCCGTGAAGGGCCGATTCAC CATCTCCAGAGACAACGCCAACACGGTGCATCTGCAAATGTACAGTCTGAAACCTGACGA

		TACGGCCGTCTATTACTGTAATTTAAATCCGCCGGGGATTCTATACTGGGGCCAGGGGAC CCAGGTCACCGTCTCCTCAGCGGCCGCTGGTAAGCCAATCCCAAACC
VarE	CDR3 93NLDPPGILY->93NLDPPGNLY	TGGTGGCGGAGGTTCTGCGGCCCAGCCGGCCATGGCACAGGTGCAGCTCGTGGAGTCGGG CGGAGGCTTGGTGCAGCCTGGGGGGTCTCTGAGACTCTCCTGTGCAGCCTCTGGAAGCTT CTTCAGCAGATATCGCATGGGCTGGTACCGCCAGGCTCCAGGGGAGCAGCGCGAGTTGGT CGCATCTATCGCTTATGATGGTAGTACAAGCTATGCAGACCCCGTGAAGGGCCGATTCAC CATCTCCAGAGACAACGCCAACACGGTGCATCTGCAAATGTACAGTCTGAAACCTGACGA TACGGCCGTCTATTACTGTAATTTAGATCCGCCGGGGAATCTATACTGGGGCCAGGGGAC CCAGGTCACCGTCTCCTCAGCGGCCGCTGGTAAGCCAATCCCAAACC
VarF	CDR3 93NLDPPGILY->93NLNPPGNLY	TGGTGGCGGAGGTTCTGCGGCCCAGCCGGCCATGGCACAGGTGCAGCTCGTGGAGTCGGG CGGAGGCTTGGTGCAGCCTGGGGGGTCTCTGAGACTCTCCTGTGCAGCCTCTGGAAGCTT CTTCAGCAGATATCGCATGGGCTGGTACCGCCAGGCTCCAGGGGAGCAGCGCGAGTTGGT CGCATCTATCGCTTATGATGGTAGTACAAGCTATGCAGACCCCGTGAAGGGCCGATTCAC CATCTCCAGAGACAACGCCAACACGGTGCATCTGCAAATGTACAGTCTGAAACCTGACGA TACGGCCGTCTATTACTGTAATTTAAATCCGCCGGGGAATCTATACTGGGGCCAGGGGAC CCAGGTCACCGTCTCCTCAGCGGCCGCTGGTAAGCCAATCCCAAACC

Table MET 3: Synthetic DNA encoding variants of R1a-B6 The sequence encoding the R1a-B6 variant (bold) is flanked by sequence overlapping with pNIBS-5 plasmid after digestion with *Not*I and *Sfi*I restriction enzymes (green highlight). These overlapping regions allow the synthetic insert to recombine with the cut plasmid in the yeast cell during transformation. The region of R1a-B6 altered in the variant, the amino acid sequence change and the sequence of the synthetic insert are all displayed (altered amino acids and nucleotides coloured red).

2.3. Assembly and transformation of R1a-B6 variant library into yeast

2.3.1. Assembly reaction

The R1a-B6 variant library was assembled from 6 overlapping synthetic single stranded oligonucleotides by splicing by overlap extension PCR (soePCR). The process was optimised to produce a 418bp major product encoding a full-length VHH domain with extensions on either end to allow homologous recombination with pNIBS-5. The six oligonucleotides were named CDR1, 2 and 3 and CONST1, 2 and 3 (table MET 4a) and were assembled in the order CONST1, CDR1, CDR2, CONST2, CDR3, CONST3 based on overlap between oligonucleotides. CONST1, 2 and 3 were standard oligonucleotides ordered with PAGE purification. CDR1, 2 and 3 were synthesised with deliberate base variation at specific bases to introduce random combinations of specific mutations at desired residues. These oligonucleotides were ordered with standard desalting (base varied oligonucleotides were not available to order with PAGE purification).

The soePCR reaction mixture and thermal cycles are described in table MET 4b,c. The reaction mixture was split equally into 4 PCR tubes.

The DNA in the reaction mixture was extracted and purified using 6 QIAquick PCR Purification kits into 150µl (total) nuclease-free water. The quantity of successfully assembled product was estimated by running the purified product on an agarose gel and photographing the gel and quantifying the intensity of the 418bp band by analysis on ImageJ[148]. The image was converted to monochrome and the brightness of the 418 band was compared to a band representing a known quantity of DNA. Background brightness was subtracted from both measurements.

A)

Name of oligonucleotide	**Sense or antisense**	**Sequence** (homology with pNIBS-5 plasmid in orange, highlighted areas overlap with the same colour, bases in brackets are mutations introduced into the unmodified R1a-B6 sequence at a frequency of 8% each)
CONST1	Sense	GA TCT GGT GGC GGA GGT TCT GCG GCC CAG CCG GCC ATG GCA CAG GTC CAG CTC GTC GAG TCG GGC GGT GGG CTG GTC CAG CCT GGT GGG TCT CTG CGT CTC TCC TGT GCA GCG TC
CDR1	Antisense	CTC ACG CTG CTC GCC TGG AGC CTG ACG GTA CCA GC(G)C(T) C(G)AT(C) ACG ATA T(A)C(A/G/T)T G(C)C(A/G/T)T GAA GAA G(C)C(A)T CC(T)C AGA CGC TGC ACA GGA GAG ACG CAG
CDR2	Sense	GCT CCA GGC GAG CAG CGT GAG TTG GTC GCA A(C/G)G(C)T ATC G(A/T)C(G)G TAT G(A)A(G)T G(A)G(A)T A(G)G(A/C/T)T A(C/G)C(T)T AG(A/C)C TAT GCA GA(G)C(G) C(T)CC GT(C)C AAG GGC CGT TTC ACC ATC TCC CG
CONST2	Antisense	G ATT ACA GTA ATA GAC GGC GGT ATC GTC AGG TTT CAG ACT GTA CAT TTG CAG ATG GAC CGT GTT GGC GTT GTC ACG GGA GAT GGT GAA ACG GCC CTT G
CDR3	Sense	CCT GAC GAT ACC GCC GTC TAT TAC TGT AAT CTA G(A)A(C/G)T CCG C(G/T)C(G)T GGG AT(A)T CTT T(G)A(C)C TGG GGC CAA GGT ACT CAG GTC ACC
CONST3	Antisense	AA TGG GTT TGG GAT TGG CTT ACC AGC GGC CGC TGA GGA GAC GGT GAC CTG AGT ACC TTG GCC CC
Full assembled product	Sense	GA TCT GGT GGC GGA GGT TCT GCG GCC CAG CCG GCC ATG GCA CAG GTC CAG CTC GTC GAG TCG GGC GGT GGG CTG GTC CAG CCT GGT GGG TCT CTG CGT CTC TCC TGT GCA GCG TCT GG(A)G AG(T)C(G) TTC TTC AG(A/C/T)C(G) AG(A/C/T)A(T) TAT CGT A(G)TG(C) G(A)G(C)C TGG TAC CGT CAG GCT CCA GGC GAG CAG CGT GAG TTG GTC GCA A(C/G)G(C)T ATC G(A/T)C(G)G TAT G(A)A(G)T G(A)G(A)T A(G)G(A/C/T)T A(C/G)C(T)T AG(A/C)C TAT GCA GA(G)C(G) C(T)CC GT(C)C AAG GGC CGT TTC ACC ATC TCC CGT GAC AAC GCC AAC ACG GTC CAT CTG CAA ATG TAC AGT CTG AAA CCT GAC GAT ACC GCC GTC TAT TAC TGT AAT CTA G(A)A(C/G)T CCG C(G/T)C(G)T GGG AT(A)T CTT T(G)A(C)C TGG GGC CAA GGT ACT CAG GTC ACC GTC TCC TCA GCG GCC GCT GGT AAG CCA ATC CCA AAC CCA TT

B)

Reagent	Volume (µl)	Concentration (µM)
Phusion Flash 2x mix	200	N/A
CONST1 (10µM)	10	0.25
CONST3 (10µM)	10	0.25
CDR1 (10µM)	2.5	0.0625
CDR2 (10µM)	2.5	0.0625
CDR3 (10µM)	2.5	0.0625
CONST2 (10µM)	2.5	0.0625
Nuclease-free water	170	N/A

C)

Step	Temperature (°C)	Time (s)
1	98	10
2	98	1
3	66	5
4	72	15
5	Go to step 2, 17 times	
6	72	60
7	4	∞

Table MET 4a,b,c: Oligonucleotides and reaction conditions used to assemble a library of mutagenised R1a-B6 genes A) Sequences and sequence variation of the assembly oligonucleotides in order of assembly (5' to 3'). Because of the assembly technique used, oligonucleotides alternate between sense and antisense. Oligonucleotide sequences are split into codons. Highlighted sections are complementary to the section highlighted in the same colour on the neighbouring oligonucleotide. At the bottom of the table, the fully assembled product is shown with all the overlap regions highlighted in their respective colours. Nucleotides in blue in the CDR oligonucleotides are variable and are followed by red nucleotides in brackets. Each base in red is present at a frequency of 8%. Bases in orange overlap with double digested pNIBS-5 plasmid allowing homologous recombination in yeast and Gibson assembly for transformation into *E.coli*. **B)** Reaction mixture used to assemble oligonucleotides into full length mutagenised R1a-B6 genes. **C)** Thermocycling conditions used in assembly reaction.

2.3.2. Pull through/mutagenesis reaction

A pull through/mutagenesis reaction followed to amplify correctly assembled product using primers Lib Fwd and Lib Rev (See table MET 5a). The primers in the pull through PCR reaction are complimentary to the 5' end of CONST1 and

the 3’ end of CONST3 so correctly assembled full VHH sequences will be amplified and truncated side products will be diluted. Conducting the pull through PCR reaction with an error prone polymerase introduces random mutations into the VHH genes.

The pull through reaction used the GeneMorph II Random Mutagenesis Kit (Agilent 200550). The reaction mixture and thermal cycle conditions are described in (table MET 5b,c).

The pull-through/mutagenesis reaction was purified using 8 QIAquick PCR Purification kits into 200µl (total) nuclease-free water.

A)

Primer Name	**Sequence**
Lib Fwd	GATCTGGTGGCGGAGGTTCTGCG
Lib Rev	AATGGGTTTGGGATTGGCTTACCAGCG

B)

Reagent	**Volume (µl)**
10x Mutazyme II reaction buffer	80
40mM dNTP mix	16
Lib Fwd primer (10µM)	25.6
Lib Rev primer (10µM)	25.6
Purified soePCR product	10 (800ng target DNA)
Mutazyme II DNA polymerase	16
Nuclease-free water	627 (to 800µl total volume)

C)

Step	Temperature (°C)	Time (s)
1	95	120
2	95	30
3	57	30
4	72	60
5	Go to step 2, 29 times	
6	72	600
7	4	∞

Table MET 5a,b,c: Pull through/mutagenesis PCR conditions and primers A) Primers used in pull through/mutagenesis PCR reaction. **B)** Pull through PCR reaction mixture. **C)** Thermocycling conditions used for pull through/mutagenesis reaction

2.3.3. Pull through PCR without mutagenesis

For purposes of testing and troubleshooting the library assembly process, it was necessary to perform the pull through amplification reaction without EP PCR. The Phusion flash mix was used instead of the Genemorph II reagents.

The assembly reaction used the same reagents, concentrations and thermal cycles but was scaled down to 16µl total volume. The pull through PCR reaction mixture and thermocycler conditions are described in tables MET6a,b.

A)

Reagent	Volume (µl)
Phusion Flash 2x mix	10
SOE PCR product (unpurified)	2
Lib Fwd primer (10µM)	1
Lib Rev primer (10µM)	1
Nuclease-free water	6

B)

Step	Temperature (°C)	Time (s)
1	98	10
2	98	1
3	72	20
4	Go to step 2, 4 times	
5	72	60
6	4	∞

Table MET 6a,b: Pull through amplification reaction without epPCR A) Reaction mixture **B)** Thermocycling conditions

2.3.4. Trial library transformation into *E.coli*

For easier and more rapid sequencing, trial libraries used to test and troubleshoot the assembly process were transformed into *E.coli*, not yeast.

Assembled and pulled-through product was purified and combined with pNIBS-5 backbone using the NEB Gibson assembly kit (NEB E5510) according to manufacturer instructions. For each trial transformation, 0.1pMol of backbone was used and 0.3pMol of insert was used in a 20 µl reaction. 6.6µl of the Gibson assembly reaction was used to transform NEB5α bacteria and 50µl of the transformation reaction was plated.

20 colonies were picked and sequenced for each trial library.

2.3.5. Transformation of library into yeast

The transformation protocol was based on the high efficiency protocol from the Sigma yeast transformation kit (not the quick protocol).

EBY100 yeast was inoculated into 10ml of YPD medium and grown to stationary phase overnight. The overnight culture was used to inoculate 150ml of YPD medium to an OD600 of 0.3 which was incubated and shaken at 30°C until the OD600 measured 1.4. It was then pelleted by centrifugation, resuspended in

150ml sterile water, re-pelleted and finally resuspended in 3ml transformation buffer (Sigma T0809) and refrigerated at 4°C.

Sixteen transformation reactions were set up and to each, 10µl of freshly boiled salmon testes DNA (Sigma D9156), 0.75µg of pNIBS-5 backbone, 2.8µg of assembled library insert and 100µl of yeast in transformation buffer were added. In addition to these 16 reactions, a reaction was set up with no insert. All reactions were vortexed. 600µl plate buffer (Sigma P8966) and 70µl of DMSO was added to each before another vortexing. The reactions were shaken at 30°C for 30 minutes and then heat shocked at 42°C for 45 minutes. Cells were centrifuged, the supernatant removed and resuspended in 1ml SD/CAA per reaction. The resuspended cell pellets were pooled together and thoroughly mixed.

The number of transformants was estimated by serially diluting a sample of the pooled reactions in SD/CAA and plating in duplicate 10, 0.1 and 0.001µl of the original pooled reaction in 100µl of SD/CAA. The colonies on the two 0.1µl plates were counted, giving an estimated number of transformants of 2.3×10^7. The same method verified that the number of transformants in the control reaction was negligible in comparison.

The rest of the pooled library was diluted 4-fold in SD/CAA and grown overnight. The overnight culture was spun down and resuspended in 15ml of SD/CAA before being stored in 15 aliquots with 20% glycerol at -80°C (section 2.1.5). Based on OD600 measurement, each library aliquot was estimated to contain 3.7×10^8 yeast cells, a 16-fold excess over the estimated number of transformants, meaning that >99.9% of transformants would be represented at least once in each aliquot.

2.4. Staining and flow cytometry of yeast

2.4.1. Biotinylation of haemagglutinin for flow cytometry experiments

Recombinant H2 (A/Canada/720/2005) and H9 (A/Hong Kong/1073/99) (eEnzyme IA-003W-005P and IA-0091W-005P) were biotinylated using the

Thermo Scientific™ EZ-Link™ Sulfo-NHS-Biotin and Labelling Kit (Thermo 21925) according to manufacturer instructions. In each case, 50µg of protein was used with a 20-fold excess of biotin and the product was eluted in PBS (Phosphate buffered saline - 8g/l NaCl, 0.2g/l KCl, 1.44g/l Na2HPO4 (Sigma S9763), 0.245g/l KH2PO4 (Sigma P0662)). The concentration of protein in the eluate was measured using the NanoDrop.

2.4.2. Inducing surface display of protein in yeast

2ml of SD/CAA was inoculated with a small amount of yeast from a colony on a petri dish or from a frozen glycerol stock and shaken for 24 hours at 30°C. 200µl of this culture was pelleted by centrifugation and resuspended in 2ml SG/R + CAA (SD/CAA with only 1g/l dextrose and with the addition of 20g/l galactose (Sigma G0750) and 20g/l raffinose (Sigma R0250)) and shaken at 20°C for 16 hours. Because of its high diversity, when the unsorted R1a-B6 variant library was induced, 1.5ml of frozen glycerol stock (one aliquot) was added to 10ml SD/CAA for the first growth step and 1ml was pelleted and resuspended in 10ml SG/R + CAA for the second. The temperature and timing of both larger volume growth steps was the same as with the smaller volume.

2.4.3. Yeast staining protocols

Unless otherwise specified, yeast cells were stained in 96-well vacuum filtration plates (Merck MAGVS2210). 2 x 10^5 cells were mixed with 200µl flow cytometry wash buffer (5g/l bovine serum albumin (BSA) (Sigma A9418), 2nM ethylenediaminetetraacetic acid (EDTA) (Sigma E9884) in PBS) and pipetted into each well of the vacuum filter plate. The buffer was removed by vacuum filtration and the cells were washed twice with 200µl ice cold wash buffer. Staining reagents were diluted in wash buffer, applied to the yeast cells and incubated for 30 minutes (at room temperature for the first staining step and at 4°C for subsequent steps). In between each staining step, the staining solution was removed by vacuum filtration and the cells were washed twice more with ice cold wash buffer. After the final staining step, cells were washed twice more and resuspended in 200µl ice cold wash buffer then transferred to round bottomed 96

well plates (Corning 3788) and put on ice until they were ready to be measured on the flow cytometer. Reagents used in flow cytometry are listed in table MET 7a and staining regimes used in different experiments are shown in table MET 7b and summarised in cartoon form in figure MET 3.

For cell sorting experiments, more than 2 x 10^5 cells needed to be stained to sufficiently cover the diversity of the pre-sort populations. In epitope mapping cell sorting, 4 x 10^5 cells were used per vacuum filter plate well and the resuspended cells from 5 identical wells were pooled to form a stained population of 2 x 10^6 cells in 1ml wash buffer. Antibody library sorting experiments used centrifugation in appropriately sized vessels to separate cells from wash and staining solutions. Volume of wash buffer and staining solutions were scaled up in proportion to number of cells but the number of washes remained the same.

A)

Reagent referred to in text	**Product name**	**Supplier**	**Product number**
Mouse α-SV5-tag	V5-Tag antibody | SV5-Pk1	Biorad	MCA1360
AF488 α-mouse	Goat anti-Mouse IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor™ 488	Invitrogen	A-11029
Biotinylated rabbit α-His-tag	Rabbit anti-6-His Tag Antibody Biotinylated	Bethyl Laboratories, Inc.	A190-114B
Streptavidin AF647 conjugate	Streptavidin, Alexa Fluor™ 647 conjugate	Invitrogen	S21374
Streptavidin AF405 conjugate	Streptavidin, Alexa Fluor™ 405 conjugate	Invitrogen	S32351
Chicken α-Myc-tag	Polyclonal chicken anti-cmyc	Bethyl Laboratories	A190-203A
α-chicken AF647 conjugate	Alexa Fluor® 647 AffiniPure Goat Anti-Chicken IgY (IgG) (H+L)	Jackson Immuno research	103-605-155
Recombinant H1 HA (A/California/07/2009)	Hemagglutinin HA (A/California/07/2009) (H1N1)	eEnzyme	IA-SW-12P

Recombinant H2 HA (A/Canada/720/2005)	Hemagglutinin HA (H2N2) (A/Canada/720/2005)	eEnzyme	IA-003W-005P
Recombinant H3 HA (A/Brisbane/10/2007)	Hemagglutinin H3 (H3N2) (A/Brisbane/10/2007)	eEnzyme	IA-0042W-005P
Recombinant H5 HA (A/Anhui/1/2005)	Hemagglutinin H5 (H5N1) (A/Anhui/1/2005)	eEnzyme	IA-0053W-005P
Recombinant H9 HA (A/Hong Kong/1073/99)	Hemagglutinin HA (H9N2) (A/Hong Kong/1073/99)	eEnzyme	IA-0091W-005P
Recombinant H9 HA (A/Guinea fowl/Hong Kong/WF10/99) biotinylated	ACROBiosystems Biotinylated Influenza A [A/Guinea fowl/Hong Kong/WF10/99(H9N2)] Hemagglutinin (HA) Protein, Avitag™,His Tag	ACROBiosystems	HA1-V82E1-25UG

B)

Staining protocol	**Experiments**	**Number of cells**	**Staining for**	**Staining reagent 1**	**Staining reagent 2**	**Staining reagent 3**
1	Alanine scan, testing of R1a-B6 clonal lineage mutants, characterisation of library samples and individual clones	2 x 10^5	-Antibody surface display -HA surface binding	25µl -Recombinant HA (12.5-50nM)	100µl -Mouse α-SV5 tag (1:1000) -Biotinylated rabbit α-his tag (1:500)	100µl -AF488 α-mouse (1:500) -Streptavidin AF647 conjugate (1:500)
2a	Library staining for H2 sorting	1 x 10^8 (first sort) 1 x 10^7 (subsequent sorts)	-Antibody surface display -HA surface binding	5ml per 10^7 cells -Mouse α-SV5 tag (1:1000)	5ml per 10^7 cells -AF488 α-mouse (1:500)	1.5ml per 10^7 cells -Recombinant H2 HA (A/Canada/720/2005) biotinylated (100nM) -Streptavidin AF647 conjugate (1µM) -stain premixed 30 minutes before use

2b	Library staining for H9 sorting (rounds 1, 2, 3a)	2 x 10^8 (first sort) 1 x 10^7 (subsequent sorts)	-Antibody surface display -HA surface binding	5ml per 10^7 cells -Mouse α-SV5-tag (1:1000)	5ml per 10^7 cells -AF488 α-mouse (1:500)	1.5ml per 10^7 cells -Recombinant H9 HA (A/Hong Kong/1073/99) biotinylated (50nM) -Streptavidin AF647 conjugate (500nM) -stain premixed 30 minutes before use
2c	Library staining for H9 sorting (Round 3b)	1 x 10^7	-Antibody surface display -HA surface binding	5ml -Mouse α-SV5-tag (1:1000)	5ml -AF488 α-mouse (1:500)	1.5ml -Recombinant H9 HA (A/Guinea fowl/Hong Kong/WF10/99) biotinylated (50nM) -Streptavidin AF647 conjugate (100nM) -stain premixed 30 minutes before use
3a	Library staining for epitope mapping experiments and characterisation of HA mutants	2 x 10^6	-HA surface display -Antibody surface binding	500µl -Recombinant VHH (100nM)	500µl -Mouse α-SV5-tag (1:500) - Chicken α-c-myc tag 1:500)	500µl -AF488 α-mouse (1:500) - α-chicken AF647 conjugate (1:500)
3b	Library staining for epitope mapping experiments (WC21 only)	2 x 10^6	-HA surface display -Antibody surface binding	500µl -Recombinant WC21 (100nM)	500µl -Mouse α-SV5-tag (1:500) -Biotinylated rabbit α-his tag (1:500)	500µl -AF488 α-mouse (1:500) -Streptavidin AF647 conjugate (1:500)

Table MET 7a,b: Reagents and staining protocols used in flow cytometry A) List of reagents used in flow cytometry yeast staining. Recombinant HA antigens used for flow cytometry were all expressed in HEK293 cells. **B)** Summary of the protocols used in this project to stain yeast expressing surface protein (VHH or HA) in preparation for flow cytometry. In all these protocols,

AF488 is used to stain for protein surface display and reagents contributing to AF488 staining are highlighted in yellow. AF647 staining in all these protocols is intended to indicate surface expressed VHH binding to HA or vice versa. Reagents contributing to AF647 staining are highlighted in blue. Protocol 1 was used for experiments in which yeast were expressing VHH but cells were not being sorted. Protocols 2a,b,c are used in VHH variant library sorts. Protocols 3a,b were used to stain yeast expressing HA for epitope mapping cell sorting and characterisation of HA mutants.

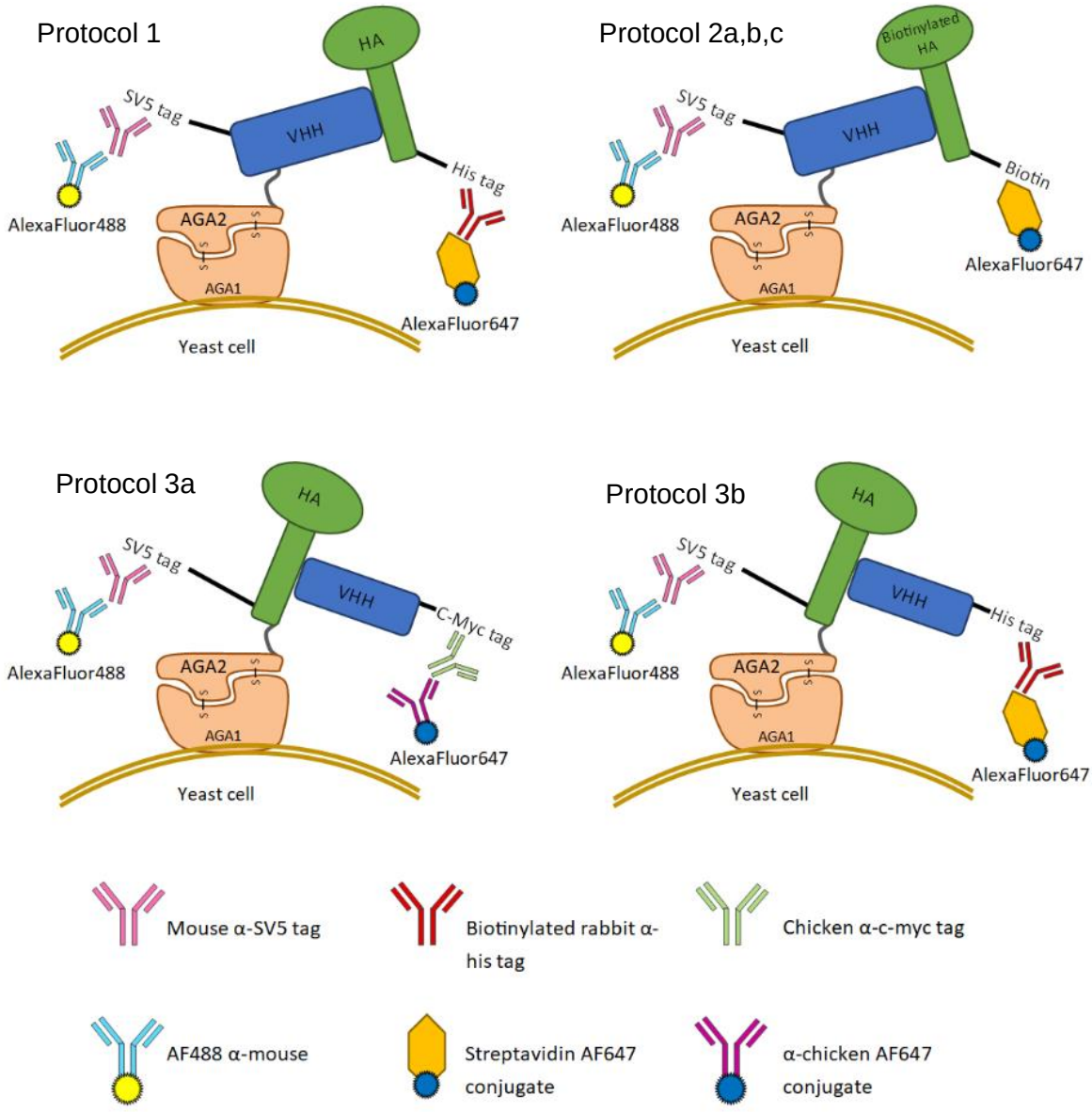


Figure MET 3: Schematics of yeast cells stained with the protocols used in this project The mechanisms of the staining protocols listed in table MET 7b are displayed in cartoon form. In each staining protocol, the level of AF488 fluorescence of a yeast cell corresponds to the level of surface protein display and the level of AF647 fluorescence corresponds to the level of HA binding to surface expressed VHH or VHH binding to surface expressed HA.

2.4.4. Compensation control populations

In experiments where yeast was expressing R1a-B6 and R1a-B6 variants, yeast expressing unmodified R1a-B6 were used in compensation controls. The AF488 control was stained according to staining protocol 1 (see table MET 7b and Figure MET 3) but without HA. The AF647 control was stained with the same protocol but with 50nM recombinant H1 HA (A/California/7/2009) and without mouse α-SV5-tag. The unstained compensation control omitted both HA and mouse α-SV5-tag.

Similarly, when HA was expressed on yeast cell surfaces, compensation populations expressed wild type H1 HA (A/California/7/2009) and were stained according to protocol 3a (see table MET 7b and figure MET 3). The AF488 control was stained omitting VHH while AF647 stained control yeast was incubated with 100nM R1a-G6[103] and omitted mouse α-SV5-tag. Unstained yeast omitted both VHH and mouse α-SV5-tag from the protocol.

2.4.5. Flow cytometry and cell sorting

Flow cytometry experiments were carried out on a BD FACSAria III machine (for cell sorting experiments) or a BD FACSCanto II machine (for all other flow cytometry experiments). Stained yeast cells were inputted into the machine either in 96 well round bottomed plates or 5ml round bottomed tubes. The forward scatter (FSC) and side scatter (SSC) of flow cytometry events was used to differentiate between single yeast cells, and cell debris or clumps of cells recorded by the machine (see figure MET 4). Events falling outside of gates defined as single yeast cells were excluded from analysis. The FITC-A, APC-A and Pacific Blue-A channels were used to measure the intensity of AF488, AF647 and AF405 staining respectively, and the voltages of these channels were adjusted at the start of every flow cytometry session to maximise the contrast between stained and unstained cells for each channel.

Flow cytometric sorting was done on a BD FACSAria III machine. Cells which fell inside the sorting gate were sorted into a 15ml V bottom tube containing 7ml SD/CAA. After sorting, the tube was capped and upended several times to

ensure cells on the tube walls were washed down into the media and the media was tipped into a baffle flask. The tube was rinsed with another 3ml SD/CAA which was also tipped into the baffle flask to maximise sorted cell recovery. The sorted cells were grown up for 48 hours to stationary phase, spun down and resuspended in 5ml SD/CAA. This medium was aliquoted into 5 tubes with 20% glycerol and frozen at -80^{o}C.

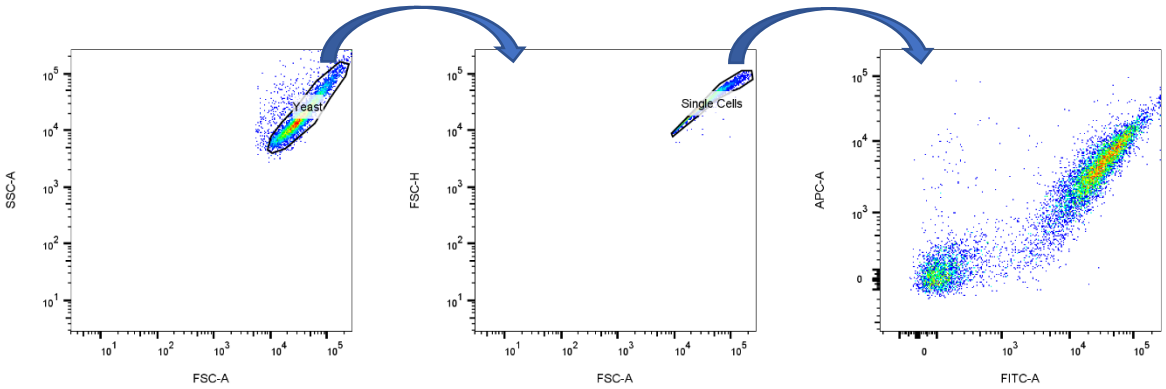


Figure MET 4: Flow cytometry gating to isolate single yeast cells Example gating of recorded flow cytometry data. The forward scatter (FSC) measures laser light deflected at a small angle by an object in the flow cytometer and this measurement correlates to cell size. The side scatter (SSC) measures light deflected at a larger angle and correlates to cell complexity. Using FSC and SSC parameters, most of the recorded events representing cell debris and clumps of yeast cells are filtered out. Two different ways of measuring FSC, FSC-A and FSC-H, are used to filter doublets of cells from the data. The remaining events are mostly single yeast cells.

2.5. Illumina sequencing and analysis of unsorted and sorted libraries

Data was collected on yeast sorting experiments with Illumina sequencing. Illumina sequencing was used both for analysis of sorting campaigns for R1a-B6 variant libraries and epitope mapping cell sorting of displayed HA variants.

For both VHH libraries and HA libraries, the unsorted library was sequenced as well as every sorted library. In addition, a monoclonal population of yeast with pNIBS-5 encoding either unmodified R1a-B6 or unmodified HA (A/California/7/2009) was sequenced so sequencing errors and mutations introduced in the gene amplification and sequencing process could be accounted for.

2.5.1. Extraction of DNA and amplification

100µl of Glycerol stock of each frozen yeast library was inoculated into 6ml of SD/CAA medium in 2 tubes and grown up to OD600 of 0.3-0.6. Libraries were pelleted by centrifugation and DNA was extracted using 4 yeast minipreps into 20µl (total) nuclease-free water.

Both VHH and epitope mapping libraries were amplified in two steps by nested PCR for maximum sensitivity. Illumina sequencing primers are listed in table MET 8a.

2.5.1.1. VHH library amplification

The region of the pNIBS-5 plasmid sequenced for all the VHH library samples included the entire VHH domain.

VHH library extracts were initially amplified in PCR reactions as described in tables MET 8b,c.

The second PCR reaction amplified DNA from the first and added Illumina sequencing adapters. This reaction is described in tables MET 8d,e. Purified DNA from this second reaction was used for Illumina sequencing.

2.5.1.2. Epitope mapping library amplification

The epitope mapping library contained two regions of interest (referred to as amplicon A and amplicon B) so, for each library, the yeast DNA extract was equally divided to amplify the amplicon A from one half and the amplicon B from the other.

Initial amplification reaction mixture and thermocycling conditions are described in tables MET 8f,g.

Second stage PCR reactions are described in tables MET 8h,i. DNA from all second stage reactions was purified using 1 Qiaquick PCR purification kit per reaction. 30µl of nuclease-free water was used to elute DNA from each reaction. Purified DNA was quantified using a NanoDrop machine and submitted for Illumina sequencing.

A)

Primer name	Region of interest	Sequence (Fwd and Rev Illumina sequencing adapter sequences in green and red)
VHH Nest Fwd	R1a-B6 residues 1-114 (whole molecule)	GTATGTTTTTGGAGGCGGAGGTTCTG
VHH Nest Rev		CAGATCAGCGGGTTTAAACGATAACAGTG
VHH Fwd		TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAGGTTCTGCGGCCCAGCCGG
VHH Rev		GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGGATTGGCTTACCAGCGGCCGC
HA AmpnA Nest Fwd	HA (A/California/7/2009) - HA1 residues 18-42	GATCTGGTGGCGGAGGTTCTG
HA AmpnA Nest Rev		CCCAAATGCAATGGGGCTACC
HA AmpnA Fwd		TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCAGCCGGCCGACACATTATGTATAG
HA AmpnA Rev		GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGCATAGTTTCCCGTTATGCTTGTCTTC
HA AmpnB Nest Fwd	HA (A/California/7/2009) - HA1 residue 291-HA2 residue 56 (95 residues)	CAGTCCACGATTGCAATACAACTTG
HA AmpnB Nest Rev		CAAATGTCCAGGAAACCATCATCAAC
HA AmpnB Fwd		TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCAAACACCCAAGGGTGCTATAAACAC
HA AmpnB Rev		GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCTGCTGTGAACTGTGTATTCATCTTTTC

B)

Reagent	Volume (µl)
Phusion Flash 2x mix	25
Yeast DNA extract	20
VHH Nest Fwd (10µM)	2.5
VHH Nest Rev (10µM)	2.5

C)

Step	Temperature (°C)	Time (s)
1	98	10
2	98	1
3	68.5	5
4	72	25
5	Go to step 2, 14 times	
6	72	60
7	4	∞

D)

Reagent	**Volume (µl)**
Phusion Flash 2x mix	25
Initial amplification reaction	5
VHH Fwd (10µM)	2.5
VHH Rev (10µM)	2.5
Nuclease-free water	15

E)

Step	**Temperature (°C)**	**Time (s)**
1	98	10
2	98	1
3	72	30
4	Go to step 2, 9 times	
5	72	60
6	4	∞

F)

Reagent (Amplicon A/B reaction)	**Volume (µl)**
Phusion Flash 2x mix	12.5
Yeast DNA extract	10
HA AmpnA/AmpnB Nest Fwd (10µM)	1.25
HA AmpnA/AmpnB Nest Rev (10µM)	1.25

G)

Step	**Temperature (°C)** (Amplicon A/B reaction)	**Time (s)**
1	98	10
2	98	1
3	67/65.5	5
4	72	25
5	Go to step 2, 14 times	
6	72	60
7	4	∞

H)

Reagent (Amplicon A/B reaction)	**Volume (µl)**
Phusion Flash 2x mix	25
Initial amplification reaction	5
HA AmpnA/AmpnB Fwd (10µM)	2.5
HA AmpnA/AmpnB Rev (10µM)	2.5
Nuclease-free water	15

I)

Step	**Temperature (°C)** (Amplicon A/B reaction)	**Time (s)**
1	98	10
2	98	1
3	67/65.5	5
4	72	25
5	Go to step 2, 9 times	
6	72	60
7	4	∞

Table MET 8a,b,c,d,e,f,g,h,i: Primers, reaction mixtures and thermocycling conditions for PCR amplification of regions of interest for Illumina sequencing Regions of interest (VHH whole gene and amplicons A and B in the HA gene) on pNIBS-5 plasmids were amplified using a nested PCR protocol from yeast DNA extract. **A)** lists primers used to amplify all three amplicons. **B)** and **C)** show the reaction mixture and thermocycling conditions for initial amplification of VHH genes. **D)** and **E)** show reaction conditions for the second amplification and addition of adapter sequences. Similarly, **F)** and **G)** describe the first amplification of amplicons A and B. Differences between the amplicon A and amplicon B protocol are coloured red (amplicon A) and green (amplicon B). **H)** and **I)** describe the second step of amplification and adapter sequence addition.

2.5.2. Illumina sequencing data processing

Illumina sequencing data was received as files of forward and reverse nucleotide reads with adapter sequences trimmed off. All data analysis was carried out in Geneious Prime version 11.0.4 (Biomatters Ltd).

Corresponding forward and reverse read files were imported into Geneious as paired end pairs of files with insert size 392bp for VHH sequencing files, 135bp for HA amplicon A and 340bp for amplicon B. Forward and reverse read files were combined with the “merge paired reads” function. Unpaired reads from this process were discarded.

2.5.2.1. VHH library epitope map data processing

For these datasets it was desirable to exclude VHH sequences with frameshifts as well as those with mutations outside the VHH encoding region. For this reason, before translation, sequences were filtered so that only merged reads starting with AGGTTCTGCGGCCCAGCCGGCC, ending with GCGGCCGCTGGTAAGCCAATCC and which were exactly 392bp long were retained and the rest were discarded. The region coding for the VHH domain was translated.

To calculate the frequency (per translated sequence) of each of the 47 targeted library mutations, the number of sequences with the correct amino acid at the correct position was queried (e.g. number of sequences with I at position 31 to find R31I frequency) then divided by the total number of translated sequences. To adjust for mutations from the sequencing process, the calculated frequencies from the monoclonal R1a-B6 control were subtracted from each of the frequencies from the other samples.

When finding frequencies of whole clones, exact matches to the full amino acid sequence were searched for in the translated sequencing data. As with individual mutations, frequencies were calculated as a proportion of the number of total sequences and, if applicable, the frequency the clone in the monoclonal R1a-B6 control sequence file was subtracted from the observed frequencies.

2.5.2.2. Epitope mapping Illumina sequencing data processing

Nucleotide sequences were translated into amino acid sequences (reading frame 1) to quantify mutations at each residue position. The number of reads with only unmodified HA (A/California/7/2009) sequence was recorded for each sample.

When tallying mutations, sequences were only counted if they had a single substitution mutation. For each of the samples, the number of substitution mutations at each position within the regions of interest (amplicons A and B, see table MET 8a) was recorded.

The same process was carried out on the control monoclonal HA sequence sample. To account for sequencing and PCR introduced mutations, the following formula was used. $M_{adj} = M_{sam} - M_{cont} (WT_{sam} / WT_{cont})$ in which M_{adj} is the adjusted number of mutations at residue position X, M_{sam} is the raw number of mutations at position X, M_{cont} is the number of mutations at position X in the control sample, WT_{sam} is the proportion of unaltered HA sequences in the sample and WT_{cont} is the proportion of unaltered HA sequences in the monoclonal control dataset.

Adjusted mutation numbers were divided by total paired reads in that sample to give a frequency of mutation at that amino acid position. To calculate enrichment factors, frequencies of mutations in sorted libraries were divided by the corresponding frequencies in the unsorted library. The enrichment factor is an estimate of the proportional change in frequency of mutation at a given position during the sorting process.

The same process was used to calculate specific mutation enrichment factors (e.g. the enrichment of mutation V40K).

2.6. Production of recombinant single domain antibodies

Single domain antibodies for this project were expressed in the periplasm of *E.coli* strains WK6 and BL21 and encoded on the plasmid pNIBS-1. TALON resin (clontech 635504) was used to bind and purify antibodies from bacterial periplasmic extract (section 2.6.3).

All liquid culture medium in this section contained 0.1µg/l carbenicillin.

2.6.1. Ligation of antibody genes into pNIBS-1 and transformation into bacteria

pNIBS-5 plasmids containing antibody genes of interest were extracted from picked yeast colonies. VHH genes were then amplified and purified (see table MET 9a). This was achieved either by PCR (using Primer 35 and Oligo 36) or transformation of yeast DNA extract into NEB5α bacteria. Plasmid was purified from grown up liquid culture of transformed bacteria.

Antibody genes were also produced as synthetic double-stranded DNA (see table MET 9b) to insert desired mutations.

In all cases, the antibody encoding insert was removed from surrounding DNA (plasmid, PCR amplicon or synthesised DNA) by *Sfi*I and *Not*I HF double restriction digestion and ligated into a pNIBS-1 backbone before being transformed into WK6 or BL21 *E.coli*.

A)

Antibody name (library of origin in brackets)	**VHH domain sequence** (differences compared to R1a-B6 sequence in red)	**DNA sequence** (starting at *Sfi*I restriction site and ending at *Not*I restriction site – both in bold)
R1a-B6	QVQLVESGGGLVQPGGSLRLSCAASG SFFSRYRMGWYRQAPGEQRELVASIA YDGSTSYADPVKGRFTISRDNANTVH LQMYSLKPDDTAVYYCNLDPPGILYW GQGTQVTVSS	**GGCCCAGCCGGCC**ATGGCACAGGTGCAGCTCGTGGAGTCGGGCGGAG GCTTGGTGCAGCCTGGGGGGTCTCTGAGACTCTCCTGTGCAGCCTCT GGAAGCTTCTTCAGCAGATATCGCATGGGCTGGTACCGCCAGGCTCC AGGGGAGCAGCGCGAGTTGGTCGCATCTATCGCTTATGATGGTAGTA CAAGCTATGCAGACCCCGTGAAGGGCCGATTCACCATCTCCAGAGAC AACGCCAACACGGTGCATCTGCAAATGTACAGTCTGAAACCTGACGA TACGGCCGTCTATTACTGTAATTTAGATCCGCCGGGGATTCTATACT GGGGCCAGGGGACCCAGGTCACCGTCTCCTCA**GCGGCCGC**
R4A (H2R4)	QVQLVESGGGLVQPGGSLRLSCAASG RFFSRYRMAWYRQAPGEQRELVASIT YDGSTNYADPVKGRFTIYRDNANTVH LQMYSLKPDDTAVYYCNLGPPGILYW GQGTQVTVSS	**GGCCCAGCCGGCC**ATGGCACAGGTCCAGCTCGTCGAGTCGGGCGGTG GGCTGGTCCAGCCTGGTGGGTCTCTGCGTCTCTCCTGTGCAGCGTCT GGGAGGTTCTTCAGCAGATATCGTATGGCCTGGTACCGTCAGGCTCC AGGCGAGCAGCGTGAGTTGGTCGCAAGTATCACGTATGATGGTAGTA CTAACTATGCAGACCCCGTCAAGGGCCGTTTCACCATCTACCGTGAC AACGCCAACACGGTCCATCTGCAAATGTACAGTCTGAAACCTGACGA TACCGCCGTCTATTACTGTAATCTAGGTCCGCCTGGGATTCTTTACT GGGGCCAAGGTACTCAGGTCACCGTCTCCTCA**GCGGCCGC**
G1 (H9R3a)	QVQLVESGGGLVQPGGSLRLSCAASG SFFSIYRMGWYRQAPGEQRELVASIA YDGGTNYADPVKGRFTISRDNANTVH LLMYSLKPDDTAVYYCNLDPPGILYW GQGTQVTVSS	**GGCCCAGCCGGCC**ATAGCACAGGTCCAGCTCGTCGAGTCGGGCGGTG GGCTGGTCCAGCCTGGTGGGTCTCTGCGTCTCTCCTGTGCAGCGTCT GGGAGCTTCTTCAGCATTTATCGTATGGGCTGGTACCGTCAGGCTCC AGGCGAGCAGCGTGAGTTGGTCGCAAGTATCGCGTATGATGGTGGTA CTAACTATGCAGACCCCGTCAAGGGCCGTTTCACCATCTCCCGTGAC AACGCCAACACGGTCCATCTGCTAATGTACAGTCTGAAACCTGACGA TACCGCCGTCTATTACTGTAATCTAGATCCGCCTGGGATTCTTTACT GGGGCCAAGGTACTCAGGTCACCGTCTCCTCA**GCGGCCGC**

G3 (H9R3a)	QVQLVESGGGLVQPGGSLRLSCAASG SFFSIYRMGWYRHAPGEQRELVASIT YDGSTNYADPVKGRFTISRDNANTVH LQMYSLKPDDTAVYYCNLDPAGILYW GQGTQVTVSS	**GGCCCAGCCGGCC**ATGGCACAGGTCCAGCTCGTCGAGTCGGGCGGTG GGCTGGTCCAGCCTGGTGGGTCTCTGCGTCTCTCCTGTGCAGCGTCT GGGAGCTTCTTCAGCATATATCGTATGGGCTGGTACCGTCATGCTCC AGGCGAGCAGCGTGAGTTGGTCGCAAGTATCACGTATGATGGTAGTA CTAACTATGCAGACCCCGTCAAGGGCCGTTTCACCATCTCCCGTGAC AACGCCAACACGGTCCATCTGCAAATGTACAGTCTGAAACCTGACGA TACCGCCGTCTATTACTGTAATCTAGATCCGGCTGGGATTCTTTACT GGGGCCAAGGTACTCAGGTCACCGTCTCCTCA**GCGGCCGC**
C23 (H9R3a)	QVQLVESGGGLVQPGGSLRLSCAASE SFFSRYRMGWYRQAPGEQRELVASIA YDGITSYADPVKGRFTISRDNANTVH LQMYSLKPDDTAVYYCNLAPPGILYW GQGTQVTVSS	**GGCCCAGCCGGCC**ATGGCACAGGTCCAGCTCGTCGAGTCGGGCGGTG GGCTGGTCCAGCCTGGTGGGTCTCTGCGTCTCTCCTGTGCAGCGTCT GAGAGCTTCTTCAGCAGATATCGTATGGGCTGGTACCGTCAGGCTCC AGGCGAGCAGCGTGAGTTGGTCGCAAGTATCGCGTATGATGGTATTA CTAGCTATGCAGACCCCGTCAAGGGCCGTTTCACCATCTCCCGTGAC AACGCCAACACGGTCCATCTGCAAATGTACAGTCTGAAACCTGACGA TACCGCCGTCTATTACTGTAATCTAGCTCCGCCTGGGATTCTTTACT GGGGCCAAGGTACTCAGGTCACCGTCTCCTCA**GCGGCCGC**
C27 (H9R3a)	QVQLVESGGGLVQPGGSLRLSCAASE SFFSTYRMGWYRQAPGEQRELIASIA YDGSTSYADPVKGRFTISRDNANTVH LQMYSLKPDDTAVYYCNLNPPGILYW GQGTQVTVSS	**GGCCCAGCCGGCC**ATGGCACAGGTCCAGCTCGTCGAGTCGGGCGGTG GGCTGGTCCAGCCTGGTGGGTCTCTGCGTCTCTCCTGTGCAGCGTCT GAGAGCTTCTTCAGCACATATCGTATGGGCTGGTACCGTCAGGCTCC AGGCGAGCAGCGTGAGTTGATCGCAAGTATCGCGTATGATGGTAGTA CTAGCTATGCAGACCCCGTCAAGGGCCGTTTCACCATCTCCCGTGAC AACGCCAACACGGTCCATCTGCAAATGTACAGTCTGAAACCTGACGA TACCGCCGTCTATTACTGTAATCTAAATCCGCCTGGGATTCTTTACT GGGGCCAAGGTACTCAGGTCACCGTCTCCTCA**GCGGCCGC**
WC4 (H9R3b)	QVQLFESGGRLVQPGGSLRLSCAASG SFFSNYRMGWYRQAPGEQRELVASIT YDGSTSYADSAKGRFTISRDNANTVH LQMYSLKPDDTAVYYCNLDPPGILYW GQGTQVTVSS	**GGCCCAGCCGGCC**ATGGCACAGGTCCAGCTCTTCGAGTCGGGCGGTA GGCTGGTCCAGCCTGGTGGGTCTCTGCGTCTCTCCTGTGCAGCGTCT GGGAGCTTCTTCAGCAATTATCGTATGGGCTGGTACCGTCAGGCTCC AGGCGAGCAGCGTGAGTTGGTCGCAAGTATCACGTATGATGGTAGTA CTAGCTATGCAGACTCCGCCAAGGGCCGTTTCACCATCTCCCGTGAC AACGCCAACACGGTCCATCTGCAAATGTACAGTCTGAAACCTGACGA TACCGCCGTCTATTACTGTAATCTAGATCCGCCAGGGATTCTTTACT GGGGCCAAGGTACTCAGGTCACCGTCTCCTCA**GCGGCCGC**
WC11 (H9R3b)	QVQLVESGGGLVQPGGSLRLSCAASG SFFSIYRMGWYRQAPGEQRELVAGIA YDGSPSYADPVKGRFTISRDNANTVH LQMYSLKPDDTAVYYCNLNPPGILDW GQGTQVTVSS	**GGCCCAGCCGGCC**ATGGCACAGGTCCAGCTCGTCGAGTCGGGCGGTG GGCTGGTCCAGCCTGGTGGGTCTCTGCGTCTCTCCTGTGCAGCGTCT GGGAGCTTCTTCAGCATATATCGTATGGGCTGGTACCGTCAGGCTCC AGGCGAGCAGCGTGAGTTGGTCGCAGGTATCGCGTATGATGGTAGTC CTAGCTATGCAGACCCCGTCAAGGGCCGTTTCACCATCTCCCGTGAC AACGCCAACACGGTCCATCTGCAAATGTACAGTCTGAAACCCGACGA TACCGCCGTCTATTACTGTAATCTAAATCCGCCTGGGATTCTTGACT GGGGCCAAGGTACTCAGGTCACCGTCTCGTCA**GCGGCCGC**
WC21 (H9R3b)	QVQLVESGGGLVQPGGSLRLSCAASE SFFSIYRMGWYRQAPGEQRELVASIA YDGTTSYADPVKGRFTISRDNANTVH LQMYSLKPDDTAVYYCNLNPPGILSW GQGTQVTVSS	**GGCCCAGCCGGCC**ATGGCACAGGTCCAGCTCGTCGAGTCGGGCGGTG GGCTGGTCCAGCCTGGTGGGTCTCTGCGTCTCTCCTGTGCAGCGTCT GAGAGCTTCTTCAGCATATATCGTATGGGCTGGTACCGTCAGGCTCC AGGCGAGCAGCGTGAGTTGGTCGCAAGTATCGCGTATGATGGTACTA CTAGCTATGCAGACCCCGTCAAGGGCCGTTTCACCATCTCCCGTGAC AACGCCAACACGGTCCATCTGCAAATGTACAGTCTGAAACCTGACGA TACCGCCGTCTATTACTGTAATCTAAATCCGCCTGGGATTCTTTCCT GGGGCCAAGGTACTCAGGTCACCGTCTCCTCA**GCGGCCGC**
WC25 (H9R3b)	QVQLVESGGGLVQPGGSLRLSCAASG SFFSRYRMGWYRQAPGGQRELVASIA YDGSTNYADPVKGRFTISRDNANTVH LQMYSLKPDDTAVYYCNLAPPGILYW GQGTQVTVSS	**GGCCCAGCCGGCC**ATGGCACAGGTCCAGCTCGTCGAGTCGGGCGGTG GGCTGGTCCAGCCTGGTGGGTCTCTGCGTCTCTCCTGTGCAGCGTCT GGGAGCTTCTTCAGCAGATATCGTATGGGCTGGTACCGTCAGGCTCC AGGCGGGCAGCGTGAGTTGGTCGCAAGTATCGCGTATGATGGTAGTA CTAACTATGCAGACCCCGTCAAGGGCCGTTTCACCATCTCCCGTGAC AACGCCAACACGGTCCATCTGCAAATGTACAGTCTGAAACCTGACGA TACCGCCGTCTATTACTGTAATCTAGCTCCGCCTGGGATTCTTTACT GGGGCCAAGGTACTCAGGTCACCGTCTCCTCA**GCGGCCGC**

B)

Antibody name	**VHH domain sequence** (differences compared to R1a-B6 sequence in red)	**DNA sequence** (*SfiI* restriction site and *NotI* restriction sites in bold)

VarD	QVQLVESGGGLVQPGGSLRLSCAASGS FFSRYRMGWYRQAPGEQRELVASIAYD GSTSYADPVKGRFTISRDNANTVHLQM YSLKPDDTAVYYCNLNPPGILYWGQGT QVTVSS	TGGTGGCGGAGGTTCTGC**GGCCCAGCCGGCC**ATGGCACAGGTGCAGC TCGTGGAGTCGGGCGGAGGCTTGGTGCAGCCTGGGGGGTCTCTGAGA CTCTCCTGTGCAGCCTCTGGAAGCTTCTTCAGCAGATATCGCATGGG CTGGTACCGCCAGGCTCCAGGGGAGCAGCGCGAGTTGGTCGCATCTA TCGCTTATGATGGTAGTACAAGCTATGCAGACCCCGTGAAGGGCCGA TTCACCATCTCCAGAGACAACGCCAACACGGTGCATCTGCAAATGTA CAGTCTGAAACCTGACGATACGGCCGTCTATTACTGTAATTTAAATC CGCCGGGGATTCTATACTGGGGCCAGGGGACCCAGGTCACCGTCTCC TCA**GCGGCCGC**TGGTAAGCCAATCCCAAACC
VarH	QVQLVESGGGLVQPGGSLRLSCAASGS FFSIYRMGWYRQAPGEQRELVASIAYD GSTSYADPVKGRFTISRDNANTVHLQM YSLKPDDTAVYYCNLDPPGILYWGQGT QVTVSS	TGGTGGCGGAGGTTCTGC**GGCCCAGCCGGCC**ATGGCACAGGTGCAGC TCGTGGAGTCGGGCGGAGGCTTGGTGCAGCCTGGGGGGTCTCTGAGA CTCTCCTGTGCAGCCTCTGGAAGCTTCTTCAGCATATATCGCATGGG CTGGTACCGCCAGGCTCCAGGGGAGCAGCGCGAGTTGGTCGCATCTA TCGCTTATGATGGTAGTACAAGCTATGCAGACCCCGTGAAGGGCCGA TTCACCATCTCCAGAGACAACGCCAACACGGTGCATCTGCAAATGTA CAGTCTGAAACCTGACGATACGGCCGTCTATTACTGTAATTTAGATC CGCCGGGGATTCTATACTGGGGCCAGGGGACCCAGGTCACCGTCTCC TCA**GCGGCCGC**TGGTAAGCCAATCCCAAACC
VarI	QVQLVESGGGLVQPGGSLRLSCAASGS FFSRYRMGWYRQAPGEQRELVASIAYD GSTSYADPVKGRFTISRDNANTVHLQM YSLKPDDTAVYYCNLDPPGILDWGQGT QVTVSS	TGGTGGCGGAGGTTCTGC**GGCCCAGCCGGCC**ATGGCACAGGTGCAGC TCGTGGAGTCGGGCGGAGGCTTGGTGCAGCCTGGGGGGTCTCTGAGA CTCTCCTGTGCAGCCTCTGGAAGCTTCTTCAGCAGATATCGCATGGG CTGGTACCGCCAGGCTCCAGGGGAGCAGCGCGAGTTGGTCGCATCTA TCGCTTATGATGGTAGTACAAGCTATGCAGACCCCGTGAAGGGCCGA TTCACCATCTCCAGAGACAACGCCAACACGGTGCATCTGCAAATGTA CAGTCTGAAACCTGACGATACGGCCGTCTATTACTGTAATTTAGATC CGCCGGGGATTCTAGACTGGGGCCAGGGGACCCAGGTCACCGTCTCC TCA**GCGGCCGC**TGGTAAGCCAATCCCAAACC
VarJ	QVQLVESGGGLVQPGGSLRLSCAASGS FFSRYRMGWYRQAPGEQRELVASIAYD GSTSYADPVKGRFTISRDNANTVHLQM YSLKPDDTAVYYCNLDPPGILSWGQGT QVTVSS	TGGTGGCGGAGGTTCTGC**GGCCCAGCCGGCC**ATGGCACAGGTGCAGC TCGTGGAGTCGGGCGGAGGCTTGGTGCAGCCTGGGGGGTCTCTGAGA CTCTCCTGTGCAGCCTCTGGAAGCTTCTTCAGCAGATATCGCATGGG CTGGTACCGCCAGGCTCCAGGGGAGCAGCGCGAGTTGGTCGCATCTA TCGCTTATGATGGTAGTACAAGCTATGCAGACCCCGTGAAGGGCCGA TTCACCATCTCCAGAGACAACGCCAACACGGTGCATCTGCAAATGTA CAGTCTGAAACCTGACGATACGGCCGTCTATTACTGTAATTTAGATC CGCCGGGGATTCTATCCTGGGGCCAGGGGACCCAGGTCACCGTCTCC TCA**GCGGCCGC**TGGTAAGCCAATCCCAAACC
VarP	QVQLVESGGGLVQPGGSLRLSCAASGS FFSIYRMGWYRQAPGEQRELVASIAYD GSTSYADPVKGRFTISRDNANTVHLQM YSLKPDDTAVYYCNLNPPGILDWGQGT QVTVSS	TGGTGGCGGAGGTTCTGC**GGCCCAGCCGGCC**ATGGCACAGGTGCAGC TCGTGGAGTCGGGCGGAGGCTTGGTGCAGCCTGGGGGGTCTCTGAGA CTCTCCTGTGCAGCCTCTGGAAGCTTCTTCAGCATATATCGCATGGG CTGGTACCGCCAGGCTCCAGGGGAGCAGCGCGAGTTGGTCGCATCTA TCGCTTATGATGGTAGTACAAGCTATGCAGACCCCGTGAAGGGCCGA TTCACCATCTCCAGAGACAACGCCAACACGGTGCATCTGCAAATGTA CAGTCTGAAACCTGACGATACGGCCGTCTATTACTGTAATTTAAATC CGCCGGGGATTCTAGACTGGGGCCAGGGGACCCAGGTCACCGTCTCC TCA**GCGGCCGC**TGGTAAGCCAATCCCAAACC
VarQ	QVQLVESGGGLVQPGGSLRLSCAASGS FFSIYRMGWYRQAPGEQRELVASIAYD GSTSYADPVKGRFTISRDNANTVHLQM YSLKPDDTAVYYCNLNPPGILSWGQGT QVTVSS	TGGTGGCGGAGGTTCTGC**GGCCCAGCCGGCC**ATGGCACAGGTGCAGC TCGTGGAGTCGGGCGGAGGCTTGGTGCAGCCTGGGGGGTCTCTGAGA CTCTCCTGTGCAGCCTCTGGAAGCTTCTTCAGCATATATCGCATGGG CTGGTACCGCCAGGCTCCAGGGGAGCAGCGCGAGTTGGTCGCATCTA TCGCTTATGATGGTAGTACAAGCTATGCAGACCCCGTGAAGGGCCGA TTCACCATCTCCAGAGACAACGCCAACACGGTGCATCTGCAAATGTA CAGTCTGAAACCTGACGATACGGCCGTCTATTACTGTAATTTAAATC CGCCGGGGATTCTATCCTGGGGCCAGGGGACCCAGGTCACCGTCTCC TCA**GCGGCCGC**TGGTAAGCCAATCCCAAACC

Table MET 9a,b: VHH genes encoding R1a-B6 variants expressed and characterised for this project R1a-B6 variant genes were transformed into WK6 or BL21 *E.coli* for protein display. Genes were either **A)** extracted from single colonies of yeast or **B)** ordered as synthetic double stranded DNA.

2.6.2. Growth and induction of antibody expression

pNIBS-1 with a VHH gene insert was transformed into E.coli (WK6 or BL21). Transformed bacteria were inoculated into 50mL 2xYT medium with 2% dextrose (w/v) and shaken overnight at 37°C.

2.6.2.1. Low dextrose method

The overnight culture was inoculated into 400ml of 2xYT medium with 0.1% (w/v) dextrose. For WK6, 2ml of starter culture was used and for BL21, 1.2ml was used.

The inoculated culture was shaken at 37°C until the culture reached an OD600 of 0.3-0.8, at which point the culture was allowed to rest at room temperature for 30 minutes. At this point, a sample was taken of the culture and IPTG (Sigma I5502) was added up to a concentration of 1mM and the culture was shaken at 28°C overnight.

In the morning, the induced culture was spun down and the pellet frozen.

Plasmid from the pre-induction sample was sequenced to verify the bacteria carried the correct VHH gene.

2.6.2.2. High dextrose method

In some of the antibody expressions, the overnight culture was inoculated into 2xYT medium with 2% dextrose (w/v). When the culture reached 0.3-0.8 OD600, bacteria were pelleted by centrifugation and resuspended in the same volume of 2xYT with 1mM IPTG and without additional dextrose. Otherwise, the protocol was as described in section 2.6.2.1.

2.6.3. TALON extraction of antibody from bacterial pellets

Pellets were thawed and resuspended in 30ml periprep (periplasmic preparation) buffer (200g/l sucrose (Sigma S0389), 200ml/l Tris HCl 1M pH 7.5 (Invitrogen 15567027), 1mM EDTA) per litre of induced culture and agitated for 1 hour at 4°C.

The suspension was spun at 16,300 x g at 4°C for 20 minutes. The supernatant was transferred to a new centrifuge tube and $MgCl_2$ solution was added to a final concentration of 10mM. The supernatant was centrifuged again (25,400 x g at 4°C for 15 minutes).

TALON resin (Clontech 635504) was washed 2-3 times in 10 volumes TALON equilibration buffer (Clontech 635651) and resuspended in 1 volume TALON equilibration buffer.

5ml per litre of induced culture of TALON resin in equilibration buffer was added to the periprep supernatant in a fresh tube and agitated for 1.5 hours at 4^{o}C. The resin was centrifuged and the periprep buffer was discarded. The resin was then washed in 10 resin volumes of equilibration buffer. The resin was again spun down and resuspended in 1 volume equilibration buffer. The resuspended resin was allowed to settle in a closed Poly-Prep® Chromatography Column (Biorad 7311550) and then drained. The resin was washed 3 more times by 3 volumes of equilibration buffer each time. Finally, the protein was eluted in 3 eluates of 1 resin volume using TALON elution buffer (Clontech 635651). The eluted protein solutions were stored at 4°C.

2.6.4. Clean-up of eluted protein

The amount of protein in eluates was roughly quantified using the NanoDrop machine. Eluates containing protein were pooled and buffer-exchanged with 2 x 5l of PBS using Slide-A-Lyzer cassettes 3.5k MWCO (Thermo 66330 (0.5-3ml) or 66110 (3-12ml)) according to manufacturer's instructions. If necessary, the protein was concentrated using Amicon Ultra-15 Centrifugal Filter Units (Merck UFC9003). The concentrated and buffer-exchanged protein solutions were finally re-quantified using the NanoDrop machine (section 2.1.6).

2.6.5. Quality control of expressed VHH samples with Coomassie stained SDS page gels and western blotting

Purified VHH samples were run on 4–20% Mini-PROTEAN® TGX™ Precast Gels (Biorad 4561093) using a Mini-PROTEAN® Tetra Cell (Bio-Rad 1658000). All samples were run at 100V alongside Precision Plus Protein Dual Color Standards (Biorad 1610374) in 10xTGS buffer (Biorad 1610732) diluted to 1x with distilled water. VHH samples expressed for this project were run alongside samples of R1a-B6 produced prior to the project as a positive control[103]. VHH samples, including the positive control, run on the same gel were all diluted in PBS to the same concentration (0.134, 0.120 or 0.050g/l) to match the lowest VHH stock concentration. 22.5µl of diluted VHH was mixed with 7.5µl of 4x Laemmli Sample Buffer (Biorad 1610747) and 0.75µl of β-mercaptoethanol (Gibco 21985023). This mixture was heated to 95°C for 5 minutes then vortexed before being loaded onto the gel. Each gel was run in duplicate resulting in two identical gels.

One of each pair of gels was covered with Coomassie staining solution (Biorad 1610436) and rocked at room temperature overnight. The gel was destained in a solution of 50% distilled water, 40% methanol and 10% glacial acetic acid by volume. 40ml of the destaining solution was applied and discarded 3 times, each time the gel was rocked in the solution for 30 minutes at room temperature. After the final destaining step, the gel was imaged in a Syngene PXi.

The protein on the second gel of each pair was transferred to a PVDF membrane (Merck IPVH00010) according to the protocol supplied with the gel. the transfer was carried out at 100V over 45 minutes.

Post transfer, the PVDF membrane was rinsed twice in distilled water and agitated overnight at 4°C in 40ml PBS plus 5% marvel dried skimmed milk powder. The membrane was rinsed twice more in distilled water before being stained in 40ml PBS with 2% milk powder and 1:2000 α-Myc-tag horseradish peroxidase antibody conjugate (Roche 11814150001) agitated at room temperature for 2 hours. The membrane was washed for 10 minutes twice in PBS plus 0.1% (v/v) TWEEN 20 (Sigma 1379) and once with PBS only. Finally, 15ml

of sigmafast tablet (Sigma D4418) staining solution was prepared according to manufacturer's instructions and the membrane was agitated in the solution for 15 minutes. The membrane was rinsed in distilled water and allowed to dry. The developed membrane was imaged in a Syngene PXi.

Stocks of recombinant VHH antibodies supplied by Simon Hufton and Chris Ball (Biotherapeutics Division, Scientific Research & Innovation, Medicines and Healthcare Products Regulatory Agency, South Mimms, Potters Bar, UK) were also used in experiments and are listed in table MET 10. These stock VHH domains included the same tags and localisation sequences as the antibodies expressed in this project. Production of these antibodies is described in [103].

Antibody name	VHH domain sequence	Reference
R1a-G6	QVQLVESGGGLVQSGGSLRLSCAASGSMSRIITMGWYRQ APGMERELVAVIGNNDNTVYGDSVQGRFTVSRDNAKNTA YLQMNSLNAEDTAMYYCKISTLTPPHEYWGQGTQVTVSS	[103]
R1a-G2	QVQLVESGGGLVQPGGSLRLSCAASGSLNSINAMGWYRQ APGKQRELVAAITNGGMTNYGGSAKGRFTISRDNAKNTV YLQMNSLKSADTAVYRCHYYSDYEGIDYWGQGTQVTVSS	[102]
R2b-E8	QVQLVESGGGLVQPGGSLRLSCAASGSFFSRYRMGWYRQ APGEQRELVAGITYDDSTNYAGSVKGRFTISRDNAKNTV HLQMNSLKPEDTAVYYCNLNPPGNLYWGQGTQVTVSS	[103]

Table MET 10: VHH domain sequences of antibodies expressed and purified prior to this project

2.7. Surface plasmon resonance (SPR) testing of antibody kinetics

All SPR[149] experiments were carried out on a Biacore T100 machine (T200 sensitivity enhanced) (GE healthcare) on Series S CM5 Sensor Chips (Cytiva BR100530) and using manufacturer supplied HBS-EP+ buffer, 10mM sodium acetate pH 5.0, 70% glycerol and 10mM glycine-HCl pH 3.0 (Cytiva BR100669, BR100351, 29207950, BR100357). Chips were coated with either recombinant HA from eEnzyme (A/California/7/2009, A/Anhui/1/2005, A/Hong Kong/1073/1999) or recombinant VHH (production described in section 2.6) using the Amine Coupling Kit (Cytiva BR100050) according to manufacturer instructions.

2.7.1. Screening of R1a-B6 variants with HA coated chips

For the HA coated chips, the target level of chip coating for HA was 9000RU and at least 8000RU of bound mass was reached for each chip. 10µl of 1µg/µl recombinant HA in PBS was mixed with 35µl HBS-EP+ buffer and used as the source of protein for chip coating. For each sensor chip used, two of the 4 cells were coated in HA and two were treated with the same coating protocol without HA added to the HBS-EP+ buffer. These two cells are referred to as blank cells.

Antibody off and on rate were determined using a single cycle kinetics (SCK) method[150]. SCK cycles were caried out without surface regeneration steps in between analytes. Instead, the dissociation time for each cycle was extended to allow the VHH to dissociate unaided. Antibody was diluted in HBS-EP+ buffer at concentrations of 5, 10, 25, 50 and 100nM and the surface of each cell on the chip was exposed sequentially to the antibody solutions in increasing order of concentration. Each solution was passed over the chip surface for 150 seconds followed by HBS-EP+ buffer for 75 seconds before the next antibody solution was introduced. After exposure to 100nM antibody, HBS-EP+ buffer was passed over the chip and dissociation was recorded for 4000 seconds.

Each HA coated cell recorded, not only signal from binding and dissociation of VHH, but also signal from changes in buffer throughout the SCK cycle. The desired binding and dissociation signal was isolated by subtracting from each HA coated cell's signal, the signal from one of the blank cells which had been subjected to the same cycle. Therefore, each SCK run resulted in two repeats produced from one HA coated cell and one blank cell each. The resultant traces represented the additional mass of VHH which was bound to the HA coated chip over the timecourse of the experiment.

The analysis software associated with the SPR machine (Biacore t200 evaluation software 3.1) estimated the interaction off and on rate from each trace by iteratively searching for a set of values matching the recorded trace (example trace shown in figure MET 5a). Where possible, each measurement was taken 3 times resulting in 6 total traces, which were averaged to produce estimated off and on rates. However, the measurable response from chips, especially in high

affinity interactions progressively declined so complete data was not gathered for every antibody tested.

2.7.2. Testing antibodies with VHH coated chips

VHH was immobilised onto SPR chips at a density of 800 to 1200RU per cell (target level 1000RU). VHH solutions used in the immobilisation protocol were diluted to 10µg/ml in sodium acetate pH 5.0. For each chip, 1 cell was coated with unmodified R1a-B6, 1 or 2 cells were coated with variants of R1a-B6 and 1 or 2 cells were left as blank cells. HA, which served as the analyte, used in these experiments was recombinant HA from eEnzyme (H1 - A/California/7/2009, H5 - A/Anhui/1/2005, H9 - A/Hong Kong/1073/1999) or from SinoBiological (H2 - A/Guiyang/1/1957, 40119-V08B, expressed in HEK293 cells).

As with HA coated chips, SCK cycles were used to estimate kinetic parameters. The H1, H5 and H9 antigen concentrations used were 5, 10, 25, 50 and 100nM. H2 antigen was used at 10, 25, 50, 100 and 200nM concentrations. Antigen was diluted in HBS-EP+ buffer. In the SCK cycles, the chips were exposed to antigen for 120s per antigen concentration and HBS-EP+ buffer was run over the chips for 75s between antigen solutions. The dissociation phase of the SCK run lasted 1000s. Chips were regenerated with a solution of 10mM glycine-HCl pH3.0 which was run over the chip for 30s. The SPR software estimated on and off rates by the same method described in section 2.7.1. An example trace is shown in figure MET 5b.

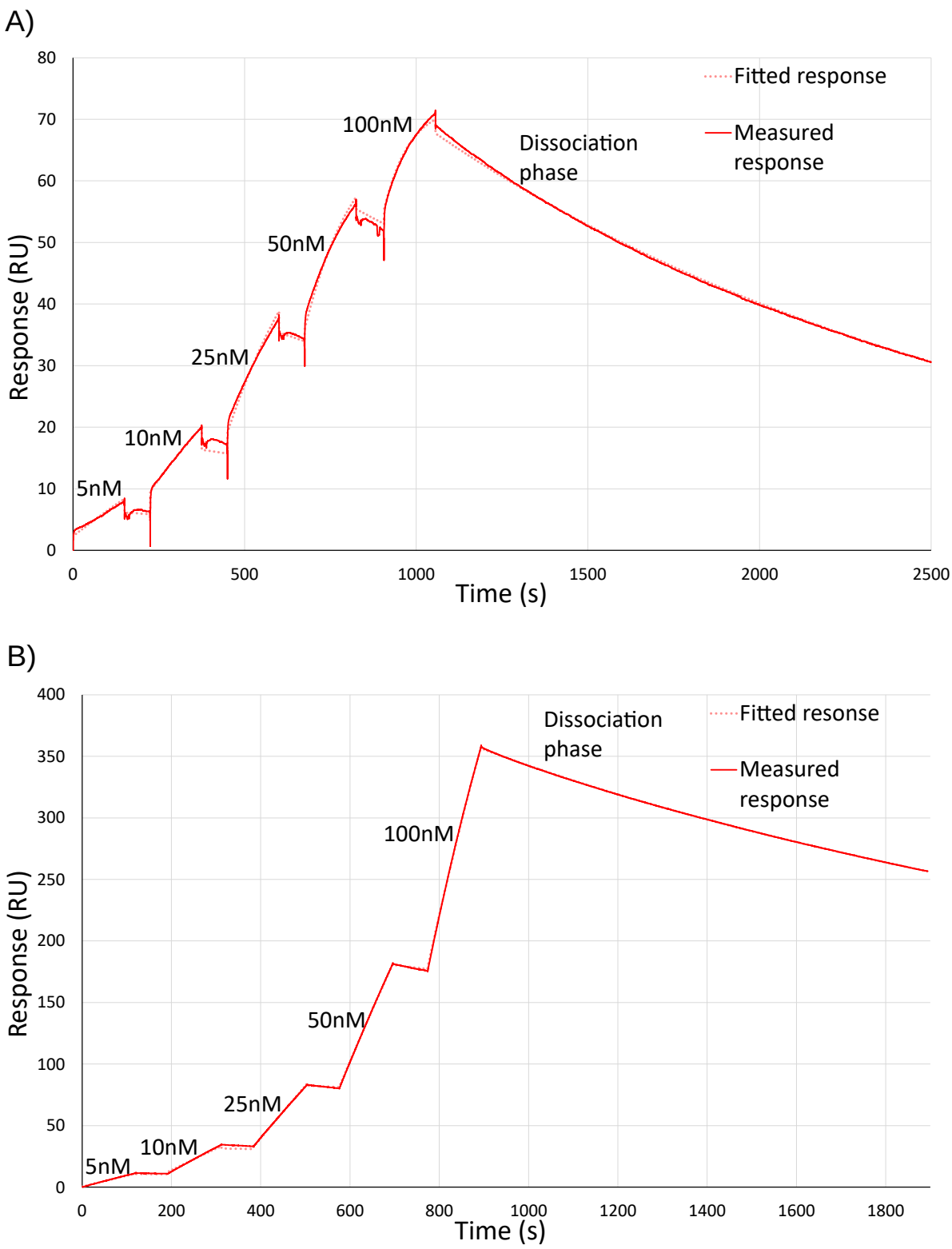


Figure MET 5a,b: example traces and fitted lines from SPR single cycle kinetics On the graphs, response, a measure of additional mass of analyte bound to an SPR surface (measured in arbitrary response units), is plotted against time. Measured response to analyte is plotted with a solid red 0line and software fitted idealised response is plotted with a dashed line. **A)** shows an example SCK trace with HA immobilised on the SPR chip and VHH as the soluble analyte. **B)** shows a trace in which VHH is immobilised on the SPR chip and solutions of HA are passed over the chip. Both traces depict the interaction of R1a-B6 and H9 HA (A/Hong Kong/1073/1999)

2.8. Pseudovirus assays

Pseudovirus (PV) used in this project was kindly provided by Joanne Del Rosario and Nigel Temperton (Viral Pseudotype Unit, Medway School of Pharmacy, The Universities of Greenwich and Kent at Medway, Chatham, UK)[151]. Pseudovirus presenting HA from the following isolates was used: H1N1 – A/England/195/2009, H2N2 – A/Korea/426/1968, H2N2 – A/England/1/1966, H5N1 - A/Vietnam/1203/2004, H7N9 – A/Shanghai/2/2013, H9N2 – A/Hong Kong/1073/1999, H9N2 – A/Chicken/Israel/2191417/2017. Creation of the PV used is described in [151]. The PV neutralisation assay protocol used in this study is based on the one used in [152] with minor modifications.

Recombinant VHH in PBS were used in this study instead of diluted sera. Assays were carried out on three separate occasions with each antibody dilution series in duplicate. The concentration range of antibodies used was 256nM to 0.5pM diluted in 20 2-fold steps (example layout of a plate in Figure MET 6). The only exception to this was H7N9 – A/Shanghai/2/2013. Replicates 2 and 3 using this PV used the VHH concentration range 1024nM - 0.5pM in 22 2-fold steps.

Each duplicate dilution series was used to produce an IC50 value using Graphpad prism 9 (GraphPad Software Inc., La Jolla, CA, USA) according to the method in [152]. Three replicate values were averaged to calculate a final IC50 value.

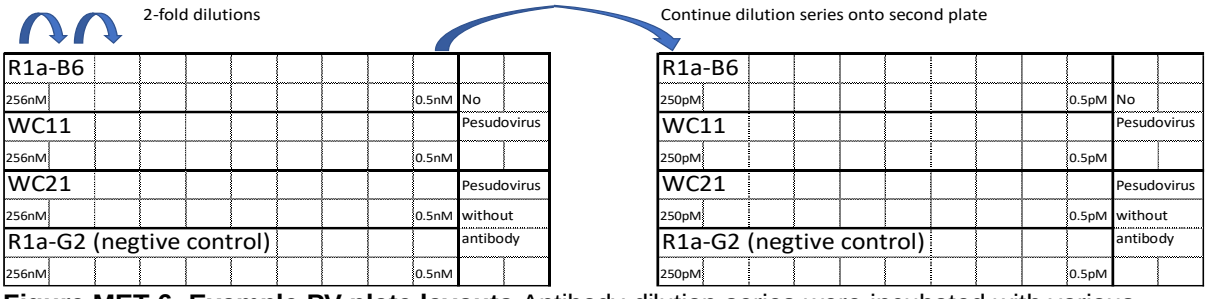


Figure MET 6: Example PV plate layouts Antibody dilution series were incubated with various PV before HEK293 cells were added. This assay tested the ability of the antibodies to inhibit the entry of the PV into cells

2.9. Enzyme-linked immunosorbent assays (ELISAs)

Indirect ELISAs[153] were used in this project to assess VHH binding to whole virus antigen standards. Representative antigen standards were used from the subtypes H1, H2, H3, H5, H7 and H9 (see table MET 11). Antigen standards, which were stored lyophilised in ampules, were reconstituted by the addition of 1ml PBS and swirled until fully dissolved. Antigen was used to coat adsorptive flat-bottomed 96-well plates (Invitrogen, 44-2404-21). Dissolved antigen was diluted 50-fold in PBS and 100µl was added to each well. Plates were incubated with antigen overnight at 4°C.

Antigen subtype	**Virus isolate**	**Product code**
H1N1	A/California/7/2009	09/174
H2N2	A/Singapore/1/1957	99/714
H3N2	A/Texas/50/2012	13/116
H5N1	A/Vietnam/1194/2004	09/184
H7N9	A/Anhui/1/2013	16/238
H9N2	A/Hong Kong/1073/99	08/208

Table MET 11 Antigen standards used in indirect ELISAs to measure VHH binding All antigen standards were sourced from NIBSC

Antigen solution was shaken out of the 96-well plates and then forcefully tapped onto paper towels to remove remaining liquid. Plates were washed twice in PBS + 0.1% TWEEN 20 then once in PBS by submersion and shaking in the wash solutions. Liquid was shaken out of plates as before.

Plates were blocked with 120µl PBS + 2% milk powder for 30 minutes with agitation and for 1 hour without, both at room temperature. As before, the blocking solution was shaken out and the plates were washed 3 times.

Duplicate dilution series rows of VHH antibody (100µl per well), diluted in PBS + 2% milk powder were added to plate columns 1-11. Antibody was diluted in √10-fold steps (a tenfold dilution every 2 steps) from 30µg/ml to 0.0003µg/ml (Example layout displayed in figure MET 7). Where possible, empty rows were positioned between rows containing different antibodies to minimise cross

contamination. Empty rows and column 12 (which served as an antibody free control) were filled with 100µl of PBS + 2% milk powder.

Plates containing VHH antibody were incubated, washed and shaken out as before.

√10-fold dilutions

R1a-B6										
30µg/ml		3		0.3		0.003		0.0003		0.00003	0
No VHH											
WC11										
30µg/ml		3		0.3		0.003		0.0003		0.00003	0
No VHH											
WC21										
30µg/ml		3		0.3		0.003		0.0003		0.00003	0

Figure MET 7: Example layout of an ELISA plate

To every well, 100µl PBS with 1:1000 α-Myc-tag horseradish peroxidase antibody conjugate and 2% milk powder was added. Plates were then agitated for 30 minutes and incubated for 30 minutes, both at room temperature. Plates were washed as before but with 4 washes with PBS + 0.1% TWEEN 20 and two washes with PBS. Plates were shaken out and remaining liquid tapped out.

The ELISA development solution was produced immediately before addition to ELISA plates and consisted of a 15:1 mixture of ELISA substrate buffer (30mM citric acid monohydrate (Sigma C1909), 0.4ml/l ProClin 300 (Sigma 48914-U), adjusted to pH 4.0 with KOH solution (Sigma P4494)) and ELISA TMB (5g/l 3,3',5,5' tetramethylbenzidine (Sigma T2885), 6ml/l 30% hydrogen peroxide solution (Sigma H1009) dissolved in a 10% acetone (VWR 200-662-2) and 90% ethanol (VWR 200-578-6) mixture(v/v)). 100µl of development solution was added to each ELISA plate well. 5 minutes later, 50µl of 1M sulfuric acid (Sigma 28-5930) was added to each well to quench the development reaction. The 450nm wavelength light absorbance of each well was measured by a Versamax

tuneable microplate reader and the associated software was used to export results.

2.9.1. ELISA data processing

ELISA data was processed using GraphPad prism 9. Duplicate pairs of datapoints were averaged and fitted to a sigmoidal curve. The EC50 (half maximal binding concentration in an ELISA) was estimated as the horizontal axis value (concentration value) at the inflection point of the sigmoidal curve

3. Results chapter 1: Characterising of R1a-B6 and library production

3.1. Chapter introduction

Technologies such as phage display[154], yeast display[155] and ribosome display[156] link an expressed protein with the gene encoding that protein. Importantly, proteins with desired phenotypes can be enriched from a heterogeneous mix of expressed proteins and their corresponding genes. Sequencing of genetic material allows identification of the enriched proteins with desired properties[157]. For example, R1a-B6 and other HA binding VHH were separated from irrelevant VHH using phage display. Capsid protein-VHH fusions were expressed on their surface of modified phage and the phage contained VHH encoding genes. Phage were selected by their ability to bind adsorbed HA on a plastic surface, thus enriching HA binding VHH genes in the phage gene pool[103].

Genotype/phenotype linkage systems allow the rapid assessment of huge numbers of proteins according to their binding properties[158-160], making these techniques ideal for *in vitro* affinity maturation of antibodies. A library of many variants of the antibody of interest can be assembled and filtered for the most improved variants compared to the unaltered antibody[161-163]. For this project, yeast cell surface display coupled with fluorescence activated cell sorting (FACS) was chosen to identify improved variants of R1a-B6[146, 164]. These techniques are expanded upon later in this chapter and in results chapter 2 (section 4).

Despite the immense throughput of yeast display and similar methods, these numbers are dwarfed by the numbers of potential variants of a given protein, even with very few substitutions[165, 166]. For example, the R1a-B6 VHH domain consists of 114 amino acids meaning that there are 2166 (19 x 114) possible single substitution variants of the protein. Adding additional substitutions causes this number to balloon rapidly. The number of R1a-B6 variants with 2 substitutions is 2.3×10^6 and the number with 3 substitutions is 1.6×10^9, meaning that a single library which includes all the triple mutant R1a-B6 variants would already be unachievable with yeast display[164, 167]. As a result, library creation strategies which constrain the number of mutations per gene copy, the number of amino acid positions mutated and/or the range of mutations at each residue

will more thoroughly search a more limited sequence space[146, 168, 169]. Conversely, less constrained combinatorial libraries will more sparsely search a much larger sequence space[170]. Of course, the success of a library which searches a limited set of mutations depends on one or more combinations of these mutations improving the function of the protein. Therefore, one approach to combinatorial library design depends on effective characterisation of the protein in question and inference of which residues and substitutions at these residues are most likely to yield functional improvement[164, 171].

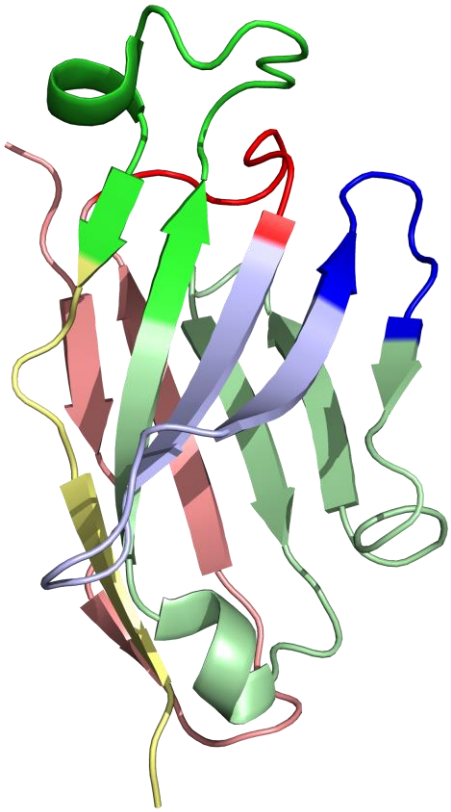
The use of a VHH heavy chain domain in this project means that despite the lack of crystal structure, the folding topology and the paratope are approximately known[105, 172]. Almost all heavy chain antibody domains bind to their epitopes with three peptide loops, known as HCDR1, HCDR2, and HCDR3[114]. Given there is no light chain associated with VHH domains, these loops will henceforth be referred to as CDR1, CDR2 and CDR3. Before, between and after these loops are the framework regions FR1, FR2, FR3 and FR4[106, 116, 173] (see figure R1 1a). The framework regions are highly conserved between antibodies and therefore usually have little influence on the specificity and affinity of the antibody, although this is a generalisation with exceptions[139, 140, 170, 174]. By contrast, the CDRs, especially the hypervariable CDR3, overwhelmingly determine affinity and specificity of the antibody, contain the most inter-antibody variation and are the site of most identified improvement mutations in *in vitro* affinity maturation studies[113, 114, 146, 164]. Concentrating library mutations in and around the CDRs, therefore, is a straightforward and effective strategy for constraining the library sequence space to include a higher proportion of potentially affinity improving mutations.

The CDRs in this project are defined as they were defined in [103]. Accordingly, CDR1 spans residues 26-33 and consists of the amino acids GSFFSRYR, CDR2 is IAYDGST (51-57) and CDR3 is NLDPPGILY (93-102) (figure R1 1b,c). All CDR designation systems have been designed around conventional antibodies. As a result, the residues defined as CDR due to their involvement with conventional antibody binding do not match precisely with the patterns found in VHH antibodies. Several residues either side of defined CDRs were found to frequently

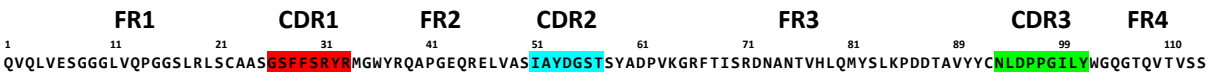
play a role in VHH binding[114]. Accordingly, some residues near to but not included in CDRs were mutated in the combinatorial library in this project.

This chapter covers the characterisation of the binding interaction of R1a-B6 with HA by both experimental and bioinformatic means. This data was used to assess the suitability of each CDR residue for mutation in an R1a-B6 combinatorial library. The process of designing, synthesising the library and transforming it into yeast is described. Finally, the library is characterised and the mutations within it are compared to the planned library design.

A)



B)



C)

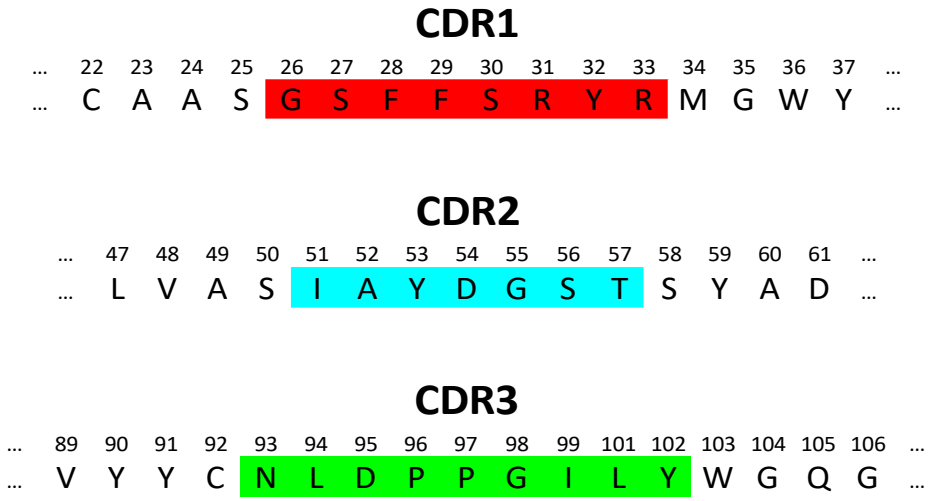


Figure R1 1a,b,c: VHH domain structure and regions A) The X-Ray crystal structure of SD38 (PDB code 6CK8[108]) exemplifies the structure of single domain antibodies. The 3 CDRs are all found on the same end of the domain and primarily determine the specificity and affinity of the antibody. Regions are coloured as follows: FR1 - pale red, CDR1 – red, FR2 – pale blue, CDR2 – blue, FR3 - pale green, CDR3 – green, FR4 - pale yellow. **B)** Sequence of R1a-B6 with regions labelled. CDRs 1, 2 and 3 are highlighted in red, green and blue respectively. **C)** CDRs of R1a-B6 as defined in [103] with the same colour scheme as **B)**.

3.2. Alanine scan of R1a-B6 CDRs

To begin to understand which residues constituted the essential parts of the paratope of R1a-B6, an alanine scan was carried out on every CDR residue[164, 171, 175]. An alanine scan examines the importance of a set of residues in a protein by testing the impact of substituting each residue in turn for alanine[176]. If one of the alanine mutants is significantly less functional (less able to bind antigen in this case) than the unmodified protein, the substituted residue is likely important to the protein's function[177]. Alanine is selected for this functional assay because the R-group on alanine, a methyl group, is small and free of polarity and charge, meaning that it is unlikely to sterically hinder or interact with the epitope in ways not found in the unmodified protein[178]. The R-group of glycine, a single hydrogen atom is smaller still than the alanine methyl group but also permits the peptide chain backbone to adopt a wider range of conformations at that position[179, 180]. Therefore, differences in protein function during a "glycine scan" could be attributable either to loss of an important R group or protein conformational distortion caused by increased backbone flexibility. This possibility is less likely with an alanine substitution. The alanine scan of R1a-B6 tested unmodified R1a-

B6 and 23 mutants (one mutant for each of the 24 CDR1, 2 and 3 residues except residue 52 which is an alanine in the unmodified protein).

Binding between R1a-B6 variants and HA was measured using a yeast display system. pNIBS-5 plasmid (see figure MET 1b) encoding R1a-B6 was modified in 23 site-directed mutagenesis reactions to yield 23 mutant plasmids encoding each of the 23 alanine scan antibody variants. The modified plasmids were transformed into EBY100 yeast. The resultant yeast colonies were sequenced to verify that they encoded the correct modified antibodies. Yeast were induced to display VHH on their surface with a galactose/raffinose containing medium and stained with recombinant HA. The staining protocol is described in section 2.4 and table MET 7b (protocol 1) and is illustrated in figure R1 2a. Briefly, yeast cells were stained with fluorophores AF488 and AF647, according to the amount of VHH displayed on the yeast cell surface and amount of recombinant HA bound to the surface expressed VHH respectively. The AF488 and AF647 fluorescence values of stained yeast populations were measured using flow cytometry. The independent measures of display and antigen binding meant that antibody binding could be detected and ranked by overlaying dot plots of different mutant VHH populations[155] (figure R1 2b). In addition, staining for surface protein display level allowed verification that the surface display profiles of mutant antibodies approximately matched the surface display profile of the unmodified R1a-B6 (figure R1 2c,e,g). This implied that differences in AF647 mean fluorescence intensity (MFI) of the stained yeast cell populations were attributable to differences in the ability of the mutants to bind antigen, not differences in antibody display levels (figure R1 2b,c,d,e,f,h). Therefore, the AF647 MFI of antibody expressing yeast cells could be used as an approximate measure of affinity relative to unmodified R1a-B6.

The alanine scanning experiment tested the binding of all 23 mutant antibodies and unmodified R1a-B6 to HA of 5 different subtypes. H1 (A/California/07/2009), H2 (A/Canada/720/2005), H5 (A/Anhui/1/2005) and H9 (A/Hong Kong/1073/1999) were selected because R1a-B6 has shown binding to these subtypes in previous experiments[103] and H3 (A/Brisbane/10/2007) was a negative control. For each combination of antibody variant and antigen, 10,000 events were recorded on the flow cytometer of the stained population and results

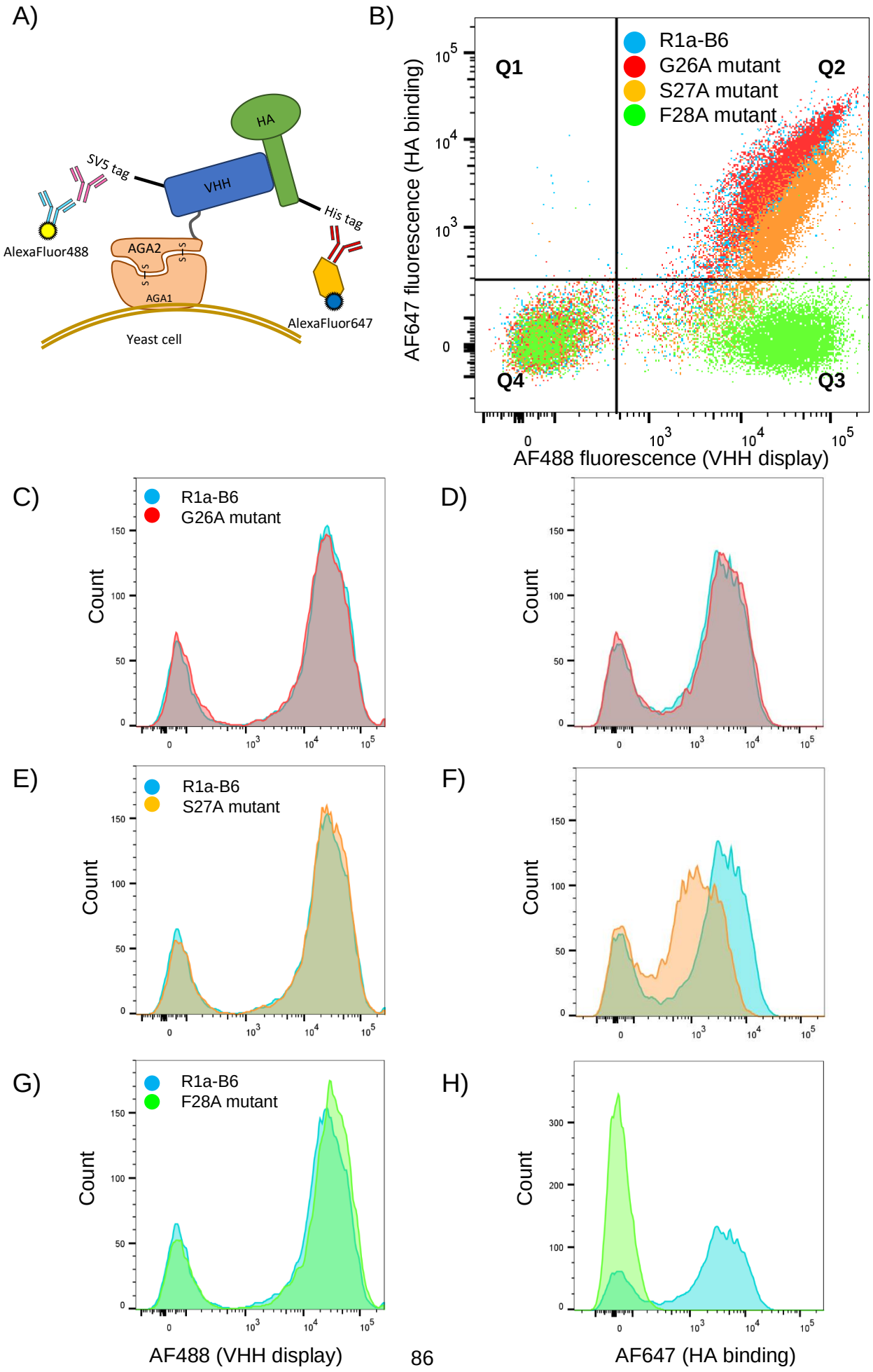
were collected in triplicate. Binding was observed by the unmodified VHH to H1, H5 and H9 but was below the limit of detection for H2 and H3. Results of these experiments are displayed in figure R1 2i.

The mutant antibodies showed a range of HA binding levels, varying from undetectable binding to binding more HA than unmodified R1a-B6 does. All three CDR loops contained residues which had no detectable effect on binding when substituted and positions that, when substituted, abolished or overwhelmingly reduced detectable binding. Key CDR residues were defined as those which, when substituted to alanine, reduced binding to H1, H5 or H9 HA by over 90%. Based on the results, CDR2 appears to play a smaller role in the antibody/antigen binding interaction compared to CDR1 and 3 as it only contains 2 key residues out of 6 residue positions substituted compared to 4 out of 8 in CDR1 and 6 out of 9 in CDR3.

Interestingly, the residues important in contributing towards H1, H5 and H9 binding appear to be noticeably different, indicating that the binding mechanisms R1a-B6 uses to bind to H1, H5 and H9 may be subtly different. Several of the substitutions which eliminate binding between the antibody and H1 or H9, including Y53A, G98A and I99A, only cause a moderate reduction in H5 binding. This phenomenon is especially pronounced in the second half of CDR3.

One substitution, R31A appears to boost the level of binding of H5 and H9 by around 2 and 3-fold respectively but reduce the binding to H1 by approximately 3-fold. Previous studies of antibody optimisation have highlighted similar substitutions, which result in an affinity trade-off, reducing affinity for one antigen while increasing affinity for another. For example, in the HA head binding antibody C05, an alanine at position 100d (Kabat) in the HCDR3 improves affinity for H3 viruses while a serine at the same position increases H1 affinity at the expense of H3 affinity[146].

In R1a-B6, another pair of substitutions, D95A and P97A, appear to improve binding to certain subtypes (H1 and H5 for D95A and H9 for P97A) without an apparent trade-off loss of affinity for another.


A)
HA
SV5 tag
VHH
His tag
AlexaFluor488
AGA2
S-S
S-S
AGA1
AlexaFluor647
Yeast cell
B)
R1a-B6
G26A mutant
S27A mutant
F28A mutant
Q1
Q2
Q3
Q4
AF647 fluorescence (HA binding)
AF488 fluorescence (VHH display)
C)
R1a-B6
G26A mutant
Count
D)
Count
E)
R1a-B6
S27A mutant
Count
F)
Count
G)
R1a-B6
F28A mutant
Count
AF488 (VHH display)
H)
Count
AF647 (HA binding)

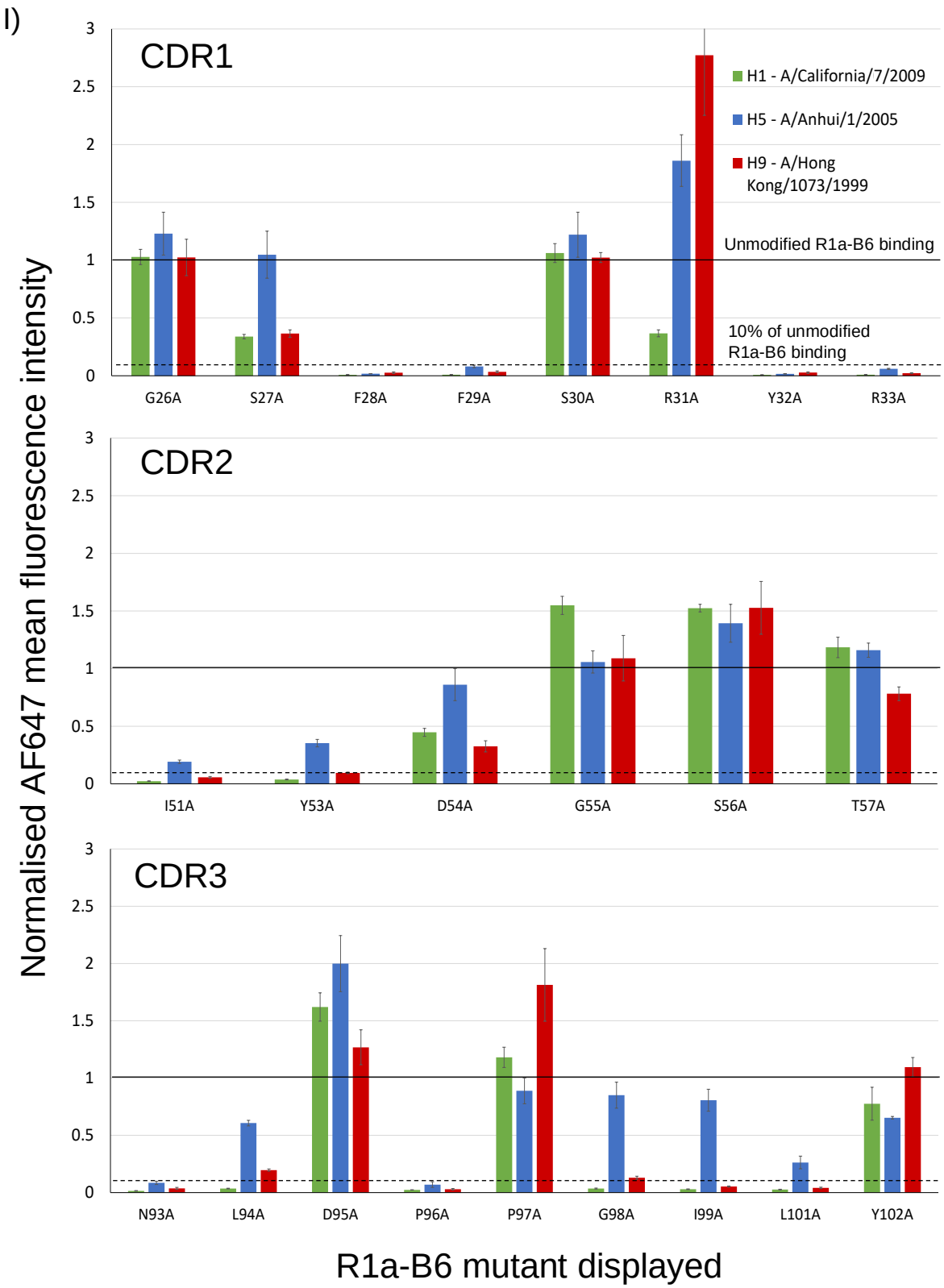


Figure R1 2a,b,c,d,e,f,g,h,i: Comparison of R1a-B6 alanine scan variants using yeast display and flow cytometry A) Cartoon depiction of a stained yeast cell displaying a single domain antibody. Staining protocol is described in section 2.4. In pink is mouse α-SV5-tag antibody, which is bound by pale blue goat anti-mouse AF488 conjugated antibody. Bound to the HA His-tag is biotinylated rabbit α-His-tag antibody (in red) which is in turn bound by streptavidin AF647 conjugate, represented as an orange hexagon. **B)** Example flow cytometric dot plot showing yeast expressing unmodified R1a-B6 (blue) compared to G26A (red), S27A (orange)

and F28A (green) antibody variants. The AF488 and AF647 florescence of a given data point depends on the number of AF488 and AF647 fluorophores bound to the corresponding yeast cell. All the yeast populations have been stained with 12.5nM H1 (A/California/07/2009) HA. The fluorescence distribution corresponding to the G26A variant mirrors that of the unmodified antibody, indicating a similar affinity for H1 HA while S27A binds markedly less and F28A has no measurable binding at all. Histograms **C)**, **E)** and **G)** illustrate that the antibody alanine scan substitutions (G26A (red), S27A (orange) and F28A (green) respectively) make no difference to levels of surface VHH display. The distributions of AF488 staining of the mutant populations are all near identical to the AF488 staining distribution of unmodified R1a-B6. On the other hand, **D)**, **F)** and **H)** show the large difference in AF647 staining distribution depending on VHH-HA affinity changes. **I)** shows the compiled data from all the alanine scan mutants stained with 12.5nM H1, H5 and H9 HA antigen. All results show the AF647 mean fluorescence intensity from the AF488 positive cells (quadrants Q2 and Q3 on **B)**) averaged from three repeats and normalised to the mean fluorescence intensity (MFI) from unmodified R1a-B6. Error bars show standard error of mean (SEM) among the 3 repeats.

3.3. Analysing the clonal lineage of R1a-B6

The *in vivo* affinity maturation process undergone by alpaca single domain antibodies is much like the equivalent process for conventional human antibodies. Millions of naïve B-cells, each displaying a randomly VDJ-recombined B-cell receptor (BCR) remain dormant until they can capture and endocytose a foreign antigen complementary to their BCR. The combined stimulus of internalising antigen and CD4+ T-cell help activates the B-cell and initiates clonal expansion. During clonal expansion, the VHH encoding genes are altered with mutations concentrated on the CDRs in a process known as somatic hypermutation[181, 182].

Following expansion, B-cells displaying mutated BCRs compete to endocytose antigen and therefore receive more T-cell help, in turn triggering more clonal expansion. Meanwhile, B-cells binding weakly to antigen receive little or no T-cell help and commit apoptosis[183, 184]. Thereby, the B-cells are under selective pressure to express the highest affinity VHH domains and the VHH expressed by the naïve B-cell will undergo many mutations during the affinity maturation process. In a mature antibody response, a single naïve B-cell activated shortly after the immune stimulus may give rise to dozens of subtly different but clonally related antibodies, all of which have been selected to bind to a single antigen[185-187].

The isolation of R1a-B6, R2b-E8 and other VHH antibodies used in this thesis is described in [103] and is summarised in figure R1 3. Briefly, an alpaca was

immunised a total of 4 times with recombinant H1 HA (A/California/7/2009) over a period of 72 days. Lymphocytes were extracted from a 10ml sample of blood taken 4 days after the last immunisation and RNA was extracted from the lymphocytes. The RNA was converted to cDNA and gene fragments encoding VHH domains were amplified from the cDNA by PCR. Amplified VHH genes were cloned into pNIBS-1 plasmid and transformed into TG1 *E.coli* to facilitate the creation of a phage antibody library.

The bacteriophages in the library expressed their VHH domains on their surface fused to a capsid protein and carried the pNIBS-1 plasmid. The library was selected for bacteriophages which bound strongly to HA resulting in selected sub-libraries with an enrichment of HA binding antibodies. The gene encoding R1a-B6 was picked from library 1a, which had undergone a single round of selection for binding to H1 (A/California/07/2009) antigen while the gene encoding R2b-E8 was picked from library 2b which consisted of library 1a after an additional round of selection for H5 (A/Vietnam/1203/2004) binders. Library 2a was also a sub library of 1a but selected for a second time on H1 (A/California/07/2009) antigen (see figure R1 3). The VHH domains encoded in all 3 selected libraries were sequenced by Illumina sequencing by Walter Ramage (Biotherapeutics Division, Scientific Research & Innovation, Medicines and Healthcare Products Regulatory Agency, South Mimms, Potters Bar, UK) using the primers listed in table R1 1.

Primer name	**Sequence** (Fwd and Rev Illumina sequencing adapter sequences in green and red)
NGSFR4	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG TGAGGAGACGGTGACCTG
NGS Alp Fr1_Q	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG CAGCCGGCCATGGCACAG

Table R1 1: Primers used for Illumina sequencing of phage display libraries

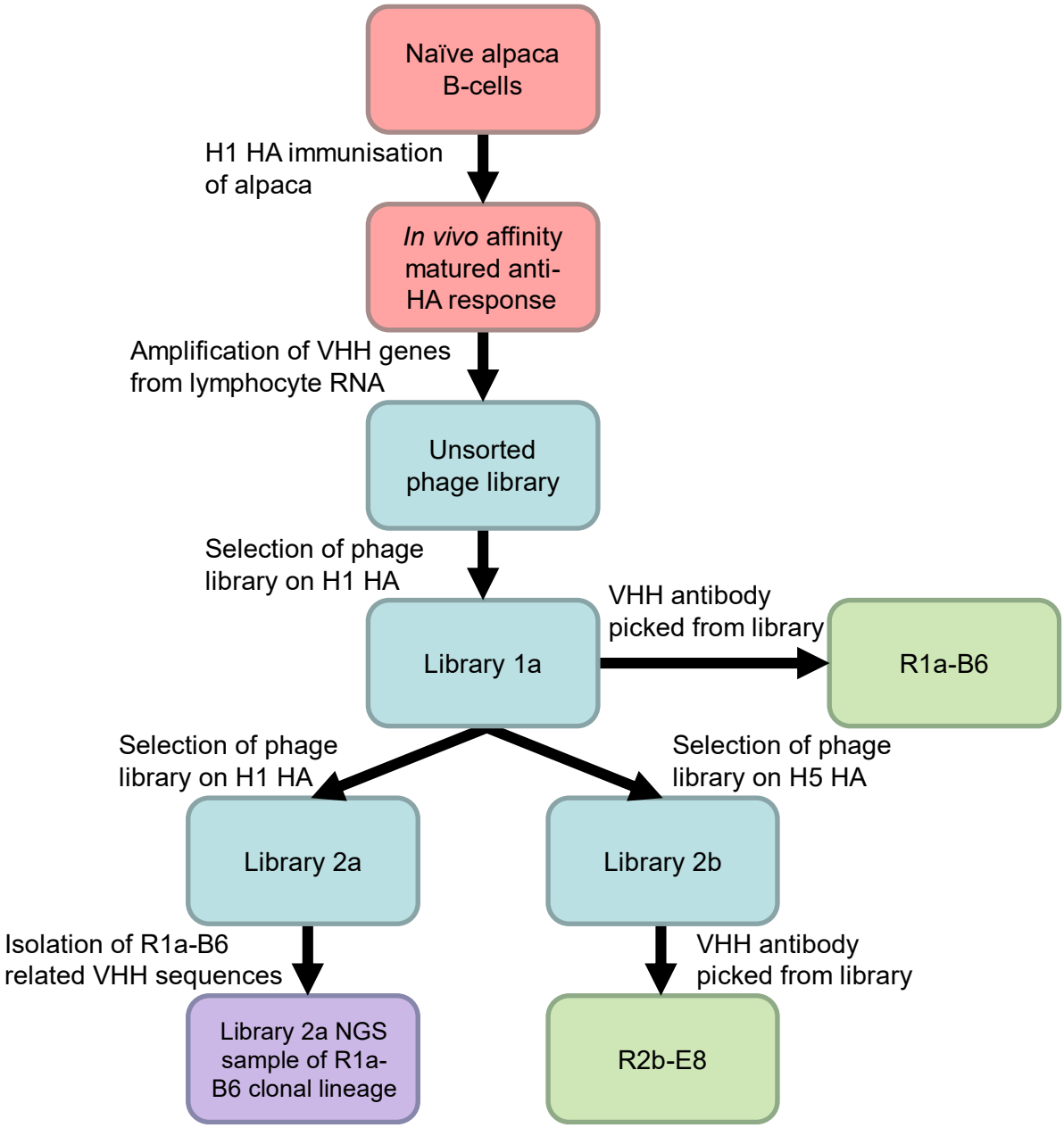


Figure R1 3: Flow chart summarising elicitation and isolation of R1a-B6 lineage antibodies and antibody sequence data Steps occurring *in vivo* are coloured in red, those carried out *in vitro* are coloured blue, monoclonal antibodies of the R1a-B6 clonal lineage are coloured green and the digital Illumina sequencing dataset used in this project is coloured purple. Illumina sequencing datasets were created from all three of the phage selected libraries (1a, 2a and 2b) but the datasets from libraries 1a and 2b were ultimately not used in the experiments of this project so are omitted from this chart. Note: though R2b-E8 was picked from library 2b, Illumina sequencing data revealed that this VHH was present in libraries 1a, 2a and 2b.

The members of a clonal lineage of antibodies can be assumed to be all adapted to bind to the same epitope[144, 185, 186, 188, 189]. Given that R1a-B6 and the clonally related VHH R2b-E8 have both been demonstrated to bind the H1 HA stem

region, it is likely that the whole lineage was affinity matured in response to the H1 antigen immunisations given to the alpaca.

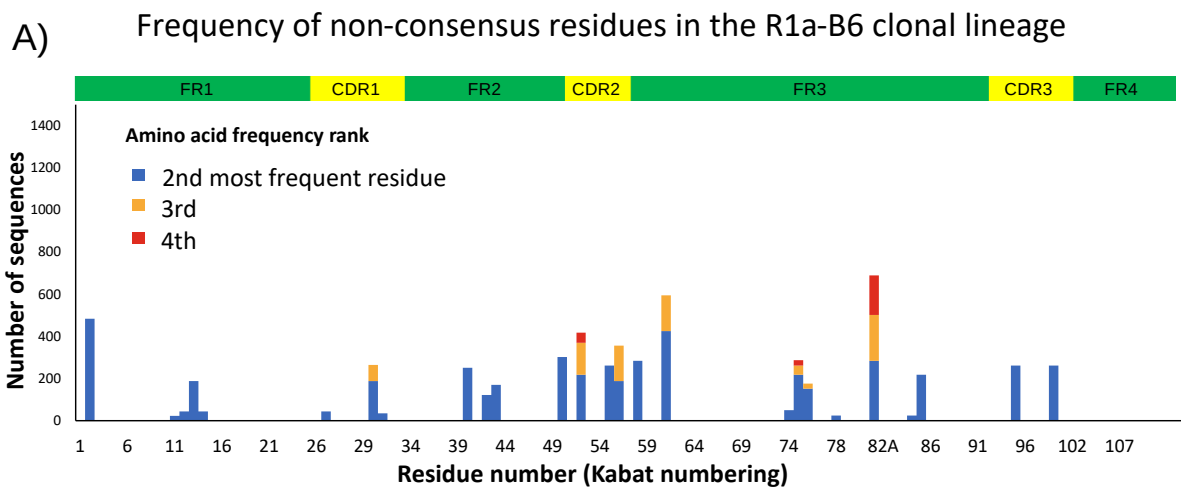
The clonal lineage of R1a-B6 was investigated by analysis of the Illumina sequence data from the phage selected libraries 1a, 2a and 2b. Illumina sequencing data analysis was carried out in Geneious Prime (Build 2019-11-07). The sequence data outputs for the 3 libraries were trimmed to include only the VHH encoding sequence and repeat sequences were concatenated into a single entry on the list of sequences. Sequences were ordered from most frequent to least frequent and only the 400 most frequent sequences from each library were retained.

R1a-B6 clonal lineage antibodies were identified in the top 400 most frequent sequences from libraries 1a, 2a and 2b. VHH domains clonally related to R1a-B6 were identified with a custom BLAST (basic local alignment search tool) search. The lists were searched with the DNA sequence encoding residues 78-113 of R1a-B6 (“CAT CTG CAA ATG TAC AGT CTG AAA CCT GAC GAT ACG GCC GTC TAT TAC TGT AAT TTA GAT CCG CCG GGG ATT CTA TAC TGG GGC CAG GGG ACC CAG GTC ACC GTC TCC TCA”) and each sequence was ranked according to its similarity to this sequence. Sequences deemed to be clonally related to R1a-B6 either encoded an identical CDR3 to R1a-B6 (NLDPPGILY) or a CDR3 identical to that of R2b-E8 (NLNPPGNLY) which differs at two residue positions from the R1a-B6 CDR3. Unrelated sequences encoded CDR3s which were much more divergent from the R1a-B6 CDR3 (for example NADSPGYDY and NSNYFGIEY) and consequently were ranked lower in the blast search. Although the CDR3 was the most characteristic feature identifying R1a-B6 clonal lineage antibodies, unrelated antibodies also diverged from R1a-B6 at CDR1 and CDR2 positions much more than R1a-B6 related antibodies. By this method, 402 reads were retrieved from library 1a, containing 10 unique sequences, 1443 reads were retrieved from library 2a with 34 unique sequences (encoding 18 VHH domains with unique peptide sequences) and 208 reads were retrieved from library 2b with 5 unique DNA sequences. Given the greater number and diversity of clonally related antibodies found in library 2a, this set of antibodies was selected for analysis. (Note: the original dataset from library 2a

included 3 additional unique VHH protein sequences which were, upon re-examination after library design and production, found to be chimeric sequences which only resembled R1a-B6 at the 5' end of the gene so were omitted from subsequent analyses).

This sample was aligned to R1a-B6, allowing the amino acid sequence of each related antibody to be compared. Using this method, the residues which vary within the sample and the frequency of variation was mapped, with particular focus on the CDRs and surrounding residues (figure R1 4a,b,c,d). With the aligned set of 1443 sequences, a consensus sequence was assembled and aligned to the set of germ line alpaca VHH V segments[190] (accession numbers in table R1 2). The closest sequence match was the segment LP VHH - s1 which is the most likely ancestral germ line sequence for the R1a-B6 clonal lineage.

Using the LP VHH - s1 segment as the root, an inferred phylogenetic tree was constructed of the R1a-B6 clonal lineage VHH genes found in library 2a (figure R1 4e). The phylogenetic tree was constructed in Geneious prime using the neighbour joining method and the Jukes Cantor genetic distance model. The mutations which occurred after the last common ancestor of this sample of sequences will be a mixture of driver and passenger mutations. Driver mutations improve the function of the antibody by improving affinity to antigen in the alpaca and therefore are selected for by the alpaca immune system. Conversely, passenger mutations make little or no difference to the affinity and are selected based on the driver mutations which happen to appear in the same sequence[185, 186].



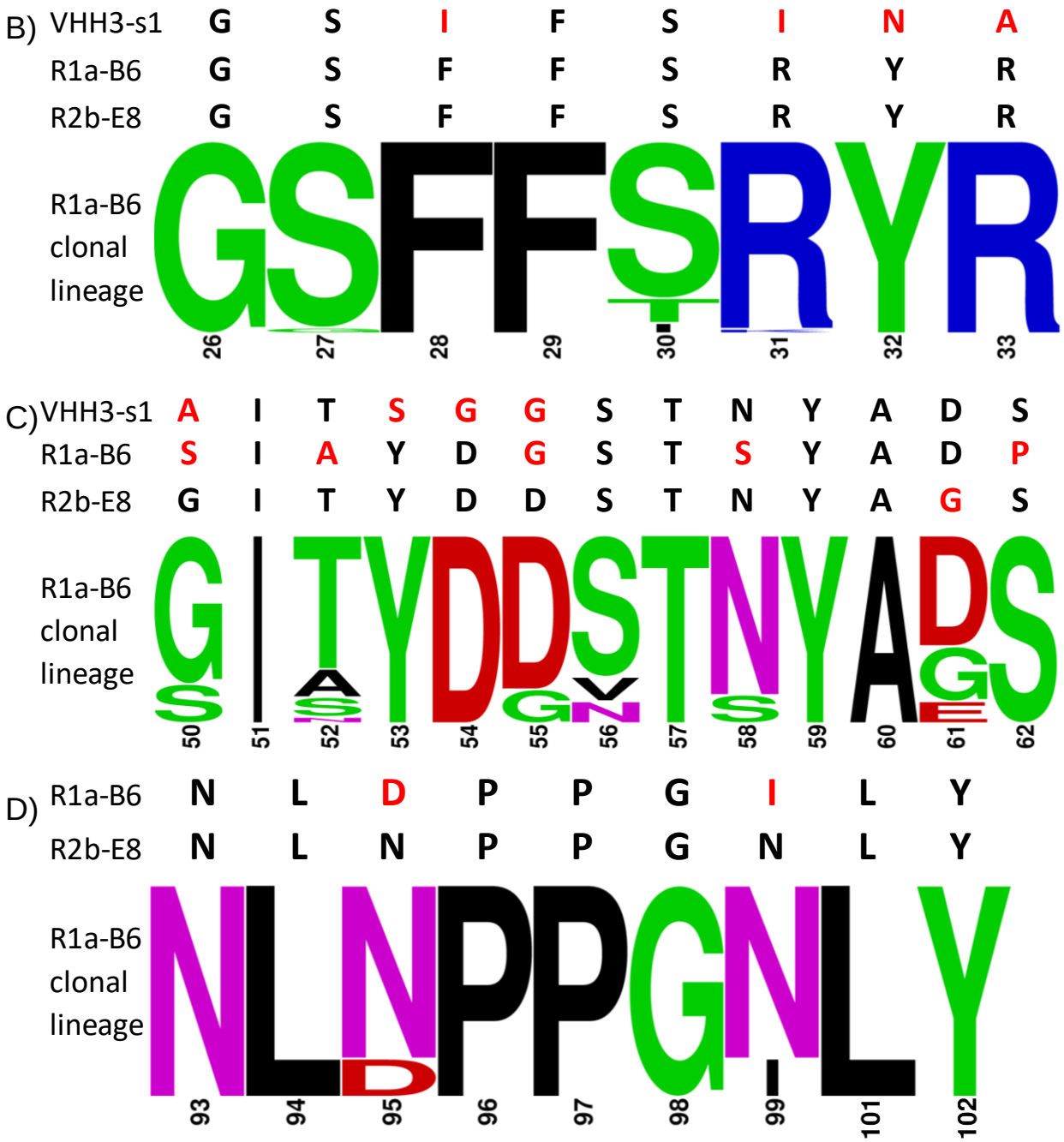

B)
VHH3-s1 G S I F S I N A
R1a-B6 G S F F S R Y R
R2b-E8 G S F F S R Y R
R1a-B6 clonal lineage
26 27 28 29 30 31 32 33
C)
VHH3-s1 A I T S G G S T N Y A D S
R1a-B6 S I A Y D G S T S Y A D P
R2b-E8 G I T Y D D S T N Y A G S
R1a-B6 clonal lineage
50 51 52 53 54 55 56 57 58 59 60 61 62
D)
R1a-B6 N L D P P G I L Y
R2b-E8 N L N P P G N L Y
R1a-B6 clonal lineage
93 94 95 96 97 98 99 101 102
E)
VHH3-s1
Lib_2a_SEQ1 (122)
R1a-B6
Lib_2a_SEQ3 (96)
Lib_2a_SEQ2 (43)
Lib_2a_SEQ7 (25)
Lib_2a_SEQ18 (24)
Lib_2a_SEQ11 (34)
Lib_2a_SEQ6 (119)
Lib_2a_SEQ8 (50)
Lib_2a_SEQ13 (22)
R2b-E8/Lib_2a_SEQ14 (240)
Lib_2a_SEQ15 (80)
Lib_2a_SEQ4 (64)
Lib_2a_SEQ5 (123)
Lib_2a_SEQ9 (143)
Lib_2a_SEQ12 (43)
Lib_2a_SEQ10 (64)
Lib_2a_SEQ16 (47)
Lib_2a_SEQ17 (104)
NLDPPGILY
NLNPPGNLY
0.02

Figure R1 4a,b,c,d,e: Analysis of the variation within the R1a-B6 clonal lineage A) depicts the location of variation within the 1443 aligned sequences clonally related to R1a-B6. These sequences were extracted from the phage library 2a Illumina sequencing data. The frequency of non-consensus amino acids at each residue position is mapped. The height of the blue, orange and red bars correspond to frequency of the 2nd, 3rd and 4th most common amino acid at each position respectively. Variation is found in every region of the molecule except FR4 but appears to be most concentrated in CDR2 and FR3. **B)**, **C)** and **D)** show sequence logos of CDR1, CDR2 (and surrounding residues) and CDR3 alongside the corresponding sequence from VHH-s1 (the putative R1a-B6 germ line sequence), R1a-B6 and R2b-E8. Amino acids that differ from clonal lineage sample consensus at that position are coloured red. Height of lettering in the sequence logos is proportional to frequency within the sample of 1443 sequences. Sequence logos were constructed on weblogo.berkeley.edu/logo.cgi[191]. **E)** depicts a phylogenetic tree constructed from the 18 VHH protein sequences found in the library 2a Illumina sequencing sample (which includes the R2b-E8 sequence), the R1a-B6 protein sequence and the alpaca VHH3 - s1 sequence. Sequences are labelled with their library of origin (library 2a), an identifying number (1-18) and in brackets, the frequency of that sequence within the clonal lineage sample. R1a-B6 and R2b-E8 are labelled in red. The VHH3-s1 sequence was defined as the root of the phylogenetic tree because it is predicted to be ancestral to the rest of the sequences. The branches encoding the two variants of the R1a-B6 CDR3 sequence are annotated. The phylogenetic tree was constructed in Geneious Prime (Build 2019-11-07) using the neighbour joining method and the Jukes Cantor genetic distance model. Sequence distance is measured in amino acid substitutions per codon. Note: the length of the root branch is not in proportion to its sequence distance from the rest of the tree.

V gene name	Subset	Accession number
LpVHH3 - s1	A	AM773548
LpVHH3 - s2	A	AM939756
LpVHH3 - s9	B	AM939763
LpVHH3 - s10	B	AM939764
LpVHH3 - s11	B	AM939765
LpVHH3 - 1	C	AM773729
LpVHH3 - s12	C	AM939752
LpVHH3 - s13	C	AM939753
LpVHH3 - s14	C	AM939754
LpVHH3 - s3	D	AM939757
LpVHH3 - s4	D	AM939758
LpVHH3 - s5	D	AM939759
LpVHH3 - s6	D	AM939760
LpVHH3 - s7	D	AM939761
LpVHH3 - s8	D	AM939762
LpVHH3 - s15	E	AM939766
LpVHH3 - s16	F	AM939755

Table R1 2: Alpaca VHH V genes accession numbers

Analysis of the position, frequency and amino acids in this sample aligned to R1a-B6 sequence can provide information on how R1a-B6 interacts with HA. Study of the evolutionary dynamics of BCR genes in clonal lineages has found that in the FRs, the B-cell affinity maturation process, on average, selects against non-synonymous mutations. Presumably mutations in these regions are unlikely to

favour high affinity antigen binding and are likely to disrupt the conserved antibody structure. CDR mutations, on the other hand, can be positively or negative selected depending on the stage of affinity maturation and the position of mutations[192, 193].

It follows therefore, that residues in the CDR which are essential to the interaction in the R1a-B6 clonal lineage are less likely to be changed in the sample as this would result in a non-functional antibody which would be selected against by the alpaca affinity maturation process[187]. As a result, residue positions that vary within the sample are less likely to be involved in interactions essential to binding because functional antibodies can be made with 2 or more different amino acids at this position.

Some agreement was found between the alanine scan data and the Illumina sequencing data. In the sample of related antibodies, the biggest hotspot of variation was in CDR2, which as noted earlier, had the fewest essential residues in the alanine scan experiment. CDR1 had less variation in the sequencing sample followed by CDR3, again mirroring the alanine scan results.

3.4. Testing the effect of clonal lineage polymorphisms on R1a-B6

It was hypothesised that the data from R1a-B6's clonal lineage could be mined for mutations which would contribute toward useful properties in R1a-B6. It is likely that other branches of the lineage possessed driver mutations which are not present in R1a-B6[194]. This principle has been demonstrated in several previous *in vitro* affinity maturation studies[144, 188, 189]. To explore this possibility, substitutions of interest were imported from other branches of the phylogenetic tree to be introduced into R1a-B6 and their effects tested using yeast display in the same way as the alanine scan mutants.

Two areas of the sequence were examined. R1a-B6, compared to the VHH - s1 germ line sequence, has a deletion in FR3 (position 74a). In many clonally related antibodies this deletion is not present and position 74a is most commonly lysine or glutamic acid. In addition, CDR3 differences within this clonal lineage were investigated because CDR3 is usually the most important CDR to the binding of

a VHH domain[113, 114, 116, 173]. The substitutions imported were two different versions of the framework region 3 sequence (VarB: 73NANT->NAKNT and VarC: 73NANT->NAENT) and two different substitutions in CDR3 separately and in combination (VarD: 93NLDPPGILY-> NLNPPGILY, VarE: 93NLDPPGILY-> NLDPPGNLY and VarF: 93NLDPPGILY-> NLNPPGNLY) (see table R1 3 and figure R1 5a,b).

R1a-B6 Variant name	**VHH domain sequence** (substitutions in red)
VarB	QVQLVESGGGLVQPGGSLRLSCAASGSFFSRYRMGWYRQAPGEQRELVASIAYDGSTSYA DPVKGRFTISRDNAKNTVHLQMYSLKPDDTAVYYCNLDPPGILYWGQGTQVTVSS
VarC	QVQLVESGGGLVQPGGSLRLSCAASGSFFSRYRMGWYRQAPGEQRELVASIAYDGSTSYA DPVKGRFTISRDNAENTVHLQMYSLKPDDTAVYYCNLDPPGILYWGQGTQVTVSS
VarD	QVQLVESGGGLVQPGGSLRLSCAASGSFFSRYRMGWYRQAPGEQRELVASIAYDGSTSYA DPVKGRFTISRDNANTVHLQMYSLKPDDTAVYYCNLNPPGILYWGQGTQVTVSS
VarE	QVQLVESGGGLVQPGGSLRLSCAASGSFFSRYRMGWYRQAPGEQRELVASIAYDGSTSYA DPVKGRFTISRDNANTVHLQMYSLKPDDTAVYYCNLDPPGNLYWGQGTQVTVSS
VarF	QVQLVESGGGLVQPGGSLRLSCAASGSFFSRYRMGWYRQAPGEQRELVASIAYDGSTSYA DPVKGRFTISRDNANTVHLQMYSLKPDDTAVYYCNLNPPGNLYWGQGTQVTVSS

Table R1 3: FR3 and CDR3 variants of R1a-B6

The effects on HA binding as measured using yeast display and flow cytometry are shown in figure R1 5c. The trend of H5 binding being resistant to substitutions in R1a-B6 continues in this dataset, variants VarB, VarC, VarE, and VarF all have significantly reduced ability to bind H1 and H9 but have similar ability as R1a-B6 to bind H5 (increased ability in the case of VarF). VarD appears to have enhanced binding to all 3 subtypes. Curiously, the VarD CDR3 sequence, NLNPPGILY, does not exist within the Illumina sequencing sample despite this sequence's apparent superiority. The only CDR3 sequences seen are NLDPPGILY and NLNPPGNLY (the CDR3 sequences of R1a-B6 and R2b-E8 respectively). The increases in HA binding for all 3 subtypes to VarD (D95N) are similar to those seen for the alanine scan mutant D95A, highlighting residue 95 as especially promising with regard to antibody optimisation.

The fact that several of these substitutions significantly reduce R1a-B6's ability to bind to H1 HA is not necessarily inconsistent with the idea that the Illumina sequencing sample contains mostly functional antibodies. The antibodies possessing these variations differ from R1a-B6 in other parts of the molecule as

well which may compensate for or synergise with differences in the FR3 and CDR3. Therefore, these variations in the context of the rest of the molecule may produce fully functional H1 binding antibodies. For example, R2b-E8 has the FR3 sequence NAKNT (as found in VarB) and the CDR3 sequence NLNPPGNLY (as found in VarF) and can bind H1 HA nearly as potently as R1a-B6[103].

A)

Kabat numbering	…	**71**	**72**	**73**	**74**	**74a**	**75**	**76**	**77**	**78**	…
R1a-B6	…	R	D	N	A	_	N	T	V	H	…
VarB	…	R	D	N	A	K	N	T	V	H	…
VarC	…	R	D	N	A	E	N	T	V	H	…

B)

Kabat numbering	…	**93**	**94**	**95**	**96**	**97**	**98**	**99**	**101**	**102**	…
R1a-B6	…	N	L	D	P	P	G	I	L	Y	…
VarD	…	N	L	N	P	P	G	I	L	Y	…
VarE	…	N	L	D	P	P	G	N	L	Y	…
VarF	…	N	L	N	P	P	G	N	L	Y	…

C)

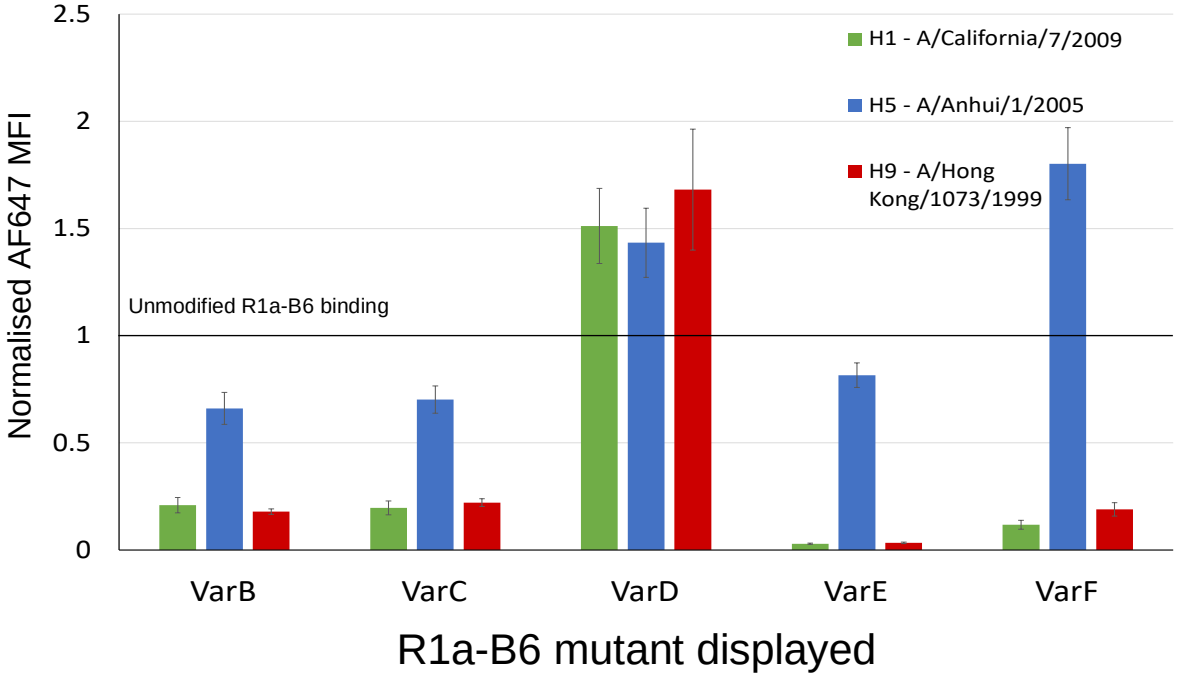


Figure R1 5a,b,c: The effect of FR3 and CDR3 variations to R1a-B6 on H1, H5 and H9 HA binding A) The protein sequences of VarB and VarC aligned to R1a-B6 around residue 74a. Relative to the germ line sequence, R1a-B6 has a deletion at position 74a. Variants VarB and

VarC revert this deletion to lysine and glutamic acid respectively, the two most common amino acids at 74a in the R1a-B6 clonal lineage sample. **B)** The CDR3 sequences of unmodified R1a-B6, VarD, VarE and VarF aligned. Variants VarD, VarE and VarF probe the CDR3 variation found in the R1a-B6 clonal lineage. VarD encodes the D95N substitution, VarE encodes I99N and VarF encodes both substitutions. **C)** Binding of VarB-F to H1, H5 and H9 HA normalised to unmodified R1a-B6. Binding was measured using yeast display under the same staining regime and conditions as the alanine scan mutants as shown in figure R1 2i. Again, results were taken in triplicate and error bars show standard error of mean.

3.5. Library design and construction

The next phase of the project was to use the data collected in sections 3.2, 3.3 and 3.4 to design a combinatorial set of mutant antibodies (a library) from which to isolate improved R1a-B6 variants. Improving properties in antibodies such as affinity or breadth of reactivity is difficult because the relationship between sequence structure and function is complex and unpredictable[195, 196] and the number of potential variations to a protein sequence is astronomical[166]. Therefore, this data was applied to predict the residues and the substitutions at those residues which stood the best chance of yielding R1a-B6 variants with a broader neutralisation profile.

3.5.1. Choice of *in vitro* antibody optimisation technology

The method chosen for high throughput library screening was yeast display. As with the alanine scanning experiment, plasmids encoding antibodies were transformed into EBY100 yeast and the yeast were stained with fluorophores for flow cytometry. However, unlike previous experiments in which each yeast population was monoclonal, the yeast in sorting experiments displayed a library of R1a-B6 variants and therefore the staining of each yeast cell depended not only on the experimental conditions and antigen used but also the antibody encoded on its pNIBS-5 plasmid. FACS allowed cells to be kept or discarded based on their levels of fluorophore staining as well as other parameters measured by the flow cytometer such as forward and side scatter. Figure R1 6a depicts the AF488 and AF647 parameter gate used for the first round H9 sorting experiment (section 4.5). A yeast cell passing through the cell sorter had to land within this gate on the dot plot to be sorted into the output tube. If the cell landed outside the gate, it would be discarded as waste. The purpose of this sorting

process was to enrich yeast cells encoding antibodies of the desired properties (in this case improved binding to H2 or H9 HA) and deplete antibodies with undesirable properties. After several cycles of selection, retaining only the most improved antibodies, the intention is that the remainder of the library is enriched for beneficial mutations and beneficial combinations of mutations (see figure R1 6a). Yeast colonies picked from repeatedly sorted libraries, therefore, should be much more likely to yield improvements in H2 or H9 binding[197, 198].

The tightly controlled multi-parameter sorting facilitated by FACS combined with yeast display is one of the key strengths of this method of *in vitro* antibody optimisation. The co-expression of VHH and SV5-tag provides independent readouts of antigen binding to VHH and VHH surface display. Consequently, antigen binding level to a yeast cell can be normalised with regard to the level of expressed VHH, allowing identification of high affinity VHH even if the display level is low[199]. By contrast, alternate antibody optimisation technologies lack this functionality. Levels of phage surface displayed antibody can be highly variable but retention of the phage on the antigen coated surface depends on the dissociation rate of phage from the surface. Library selections can therefore be biased towards highly expressed antibodies which are retained by avidity effects rather than high monovalent affinity[200]. Furthermore, yeast display allows collection of detailed data on the distribution of affinities in the library as a whole[201]. Data like this can be used to assess if the library sorting approach is working. In a successful experiment, enrichment of more strongly binding antibodies will cause the population to stain more strongly for binding to the antigen in later sorting steps.

The limiting factor for this experiment on library size was the rate of flow cytometric sorting. Cells were sorted on a BD FACS Aria III machine with a maximum sort rate of 10^4 events per second so the maximum number of cells which can be feasibly sorted in a working day is approximately 1 to 2 x 10^8 [164]. In the initial sort of a library the number of cells stained and sorted should ideally be tenfold higher than the original diversity of the library[201]. Some clones will, by chance, be underrepresented in the sample of stained cells and others will fail to express antibody so a substantial excess of stained cells will mean that the

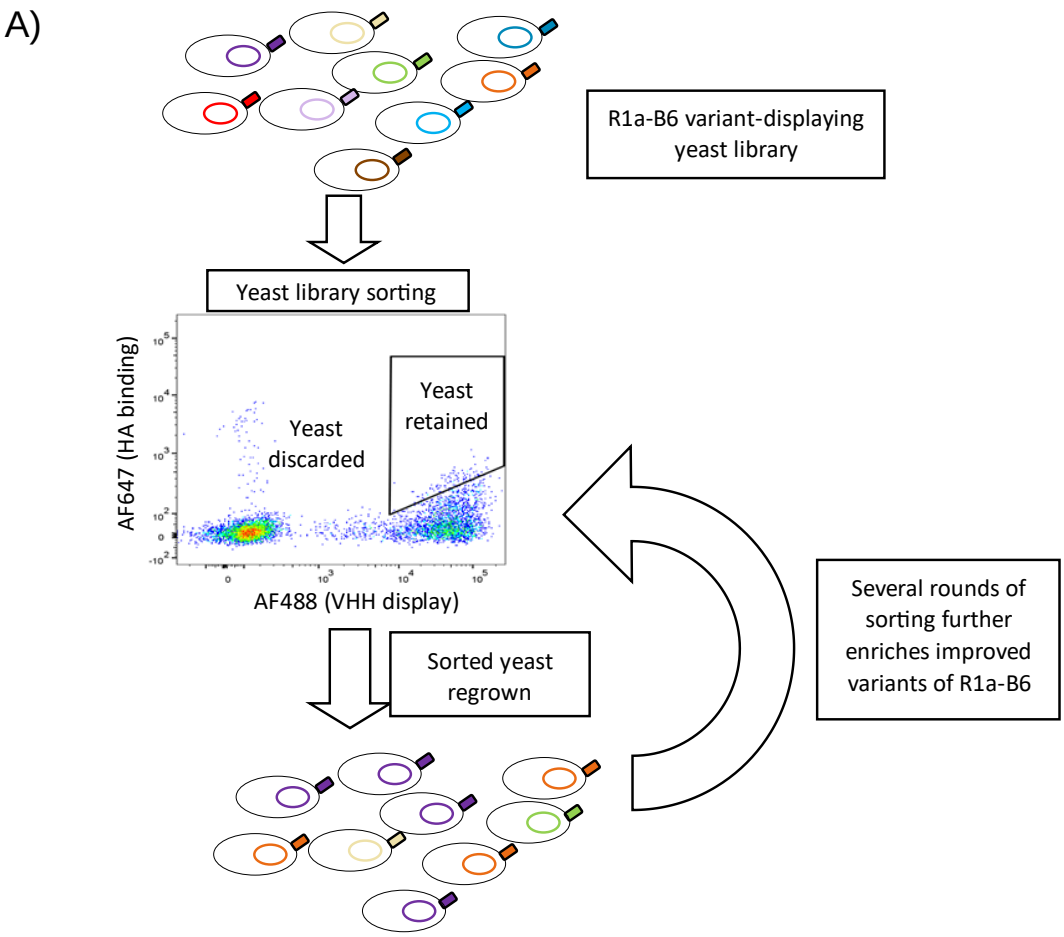
overwhelming majority of clones will be represented in the antibody expressing population more than once. These two factors limit the available sequence space to around 2 x 10^7 variants. The library was designed with these numbers in mind.

Limitations on library size can be circumvented with iterative antibody library creation strategies. For example, DNA shuffling can be used on a sorted library enriched in beneficial mutations to create a new library with reshuffled combinations of these mutations for further selection[139]. Walking mutagenesis strategies identify the most affinity improving mutations in different sections of an antibody. Like with DNA shuffling, beneficial mutations are combined for greater improvements in affinity[202]. These affinity maturation strategies search areas of sequence space already enriched with beneficial mutations, so efficiently locate variants with superior binding properties. However, the results of this project demonstrate that the diversity afforded by a single yeast library was, in this case, sufficient to identify much more broadly neutralising variants of R1a-B6.

3.5.2. Method of library assembly

The mutant R1a-B6 variant gene library was assembled by splicing by overlap extension PCR (soePCR), a method which uses overlapping single stranded DNA fragments and thermocycling to assemble larger DNA products[203]. The assembly protocol is summarised in figure R1 6b and the protocol is described in detail in section 2.3. The fragments alternate between sense and antisense. Overlapping regions of the fragments act to prime polymerisation to extend a fragment using the adjacent fragment as a template. Over several cycles, fragments are extended to full-length molecules. The two fragments at the 3' end and the 5' end (fragments CONST1 and 3) were in excess compared to the other fragments to favour the amplification of the full length VHH gene over formation of truncated products. 6 fragments were recombined including 3 fragments with variation introduced at selected codons and 3 without any mutated bases (Table MET 4a). The 3 variable fragments encode the three CDRs and the surrounding regions while the unmutated oligonucleotides encode the framework regions of the antibody.

Although the major product of the soePCR reaction is composed of the 6 oligonucleotides assembled correctly, there are substantial quantities of side products, both smaller and larger than the correct fragment size. This necessitated a second PCR (pull-through PCR) step to amplify the correct product and dilute side products. In addition, several studies have shown that although improvements in antibody affinity and function tend to be conferred by CDR mutations, important mutations for binding properties can occur outside those regions as well[139, 140, 170, 174]. The pull-through PCR step therefore was carried out using error prone PCR (epPCR) to introduce low frequency random nucleotide substitutions into the full-length antibody genes, giving the library a low level of mutation outside the CDR regions. The process of library fragment assembly is summarised in figure R1 6b, and an electrophoresis gel showing the products of the assembly reaction and pull through/mutagenesis PCR are shown in figure R1 6c.



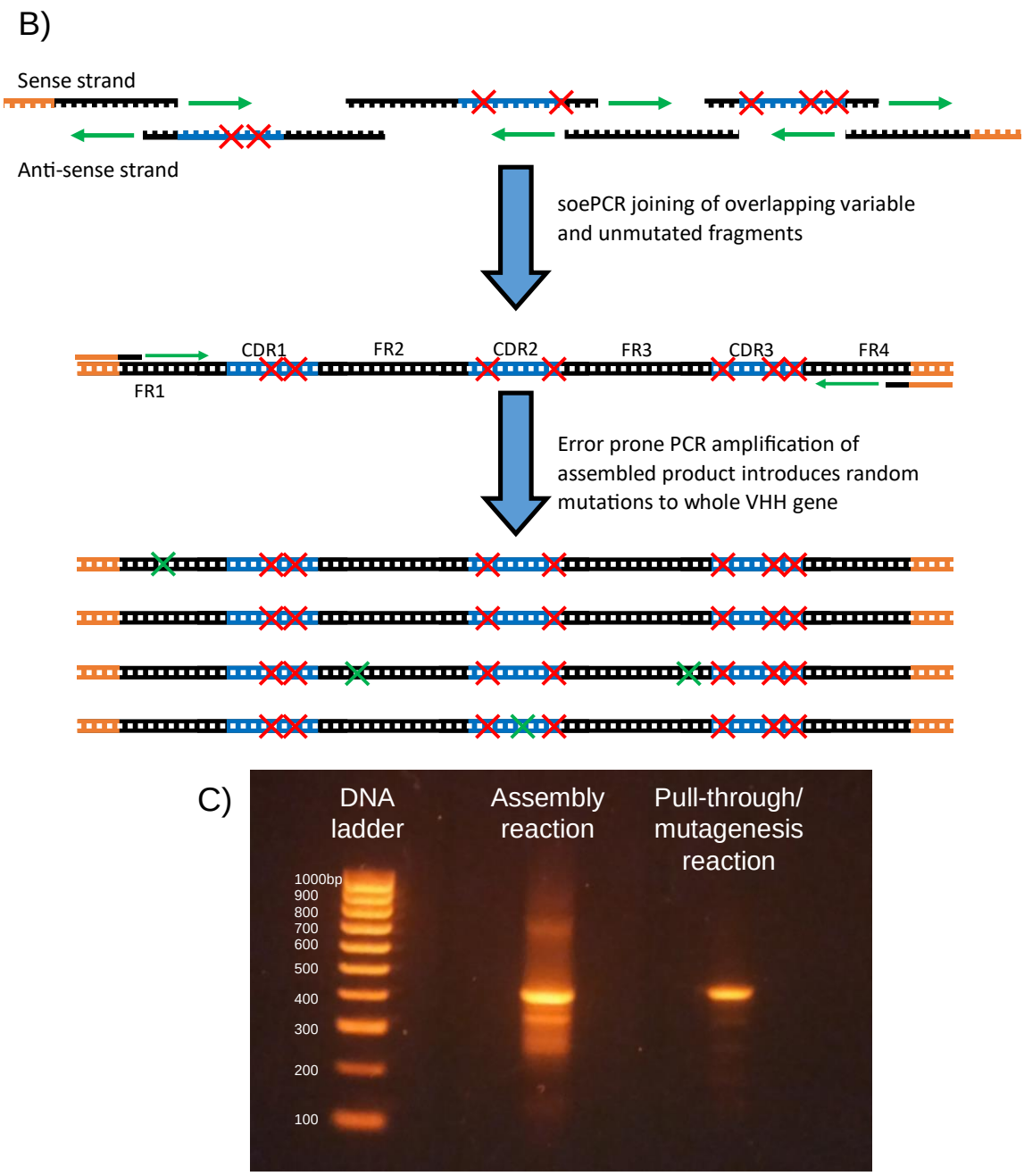


Figure R1 6a,b,c: Sorting and synthesis of R1a-B6 variant library A) Yeast sorting method to identify improved variants of R1a-B6. Repeated cycles of flow cytometric library sorting and regrowth allow the highest affinity variants to be identified from large libraries, making this technology ideal for *in vitro* antibody optimisation. **B)** Generation of a library of R1a-B6 variant genes. Targeted mutations (red crosses) were introduced into the CDRs (blue) of the R1a-B6 gene by joining variable oligonucleotides with non-mutated segments using soePCR. Random nucleotide substitutions (green crosses) were introduced by amplifying the complete gene under low fidelity conditions. The final product was transformed into EBY100 yeast alongside pNIBS-5 backbone. Overlap with the plasmid backbone (orange regions) facilitated homologous recombination forming complete plasmids inside the yeast cells[204]. **C)** Agarose gel showing the products of the assembly and pull through reactions used to produce mutagenised R1a-B6 genes for yeast display library creation (section 2.3). Comparison of band migration to the DNA ladder confirms that the major product of the assembly reaction is the correct size to be the 418bp fully assembled product of the 6 oligonucleotides present in the assembly reaction. The pull through/mutagenesis reaction amplifies the same product while diluting the side products, resulting in a strong band at ~418bp and much fainter bands elsewhere.

3.5.3. Selection of mutations

Within the CDRs (and surrounding residues), only specific residues which show some evidence of permissiveness to substitution were selected to be altered in the library. This process used data from the alanine scan experiments as well as the R1a-B6 clonal lineage sequences. Residues which, when substituted into alanine showed >10% of unmodified R1a-B6 binding for H1, H5 and H9, were mutated in the library (see figure R1 2i). These residues had demonstrated that they were inessential to R1a-B6 binding to HA. Similarly, polymorphic positions within the Illumina sequencing dataset of antibodies demonstrate amino acid plasticity at that position and, therefore, were also selected for mutation. Positions within the CDRs which fit neither of these requirements were not chosen for mutation. Mutations in residues with no evidence of amino acid plasticity (either in the alanine scan or in Illumina sequencing data) would likely have inferior binding properties compared to wild type and would take up valuable sequence space in the library which could instead be used to introduce more promising mutations or combinations of mutations.

Having selected which residues to mutate, the amino acid substitutions at these positions needed to be decided. These decisions were also based on data collected from sections 3.1 and 3.2. At each selected CDR position, the unmodified residue was chosen to be the most frequent residue, and each introduced amino acid (1-4 different amino acids) was present at a frequency of around 8%. This strategy is based on previous antibody library affinity maturation studies in which the improved variants identified were only different from the wild type by a small number of changes to the CDRs while the average number of mutations relative to wild type in the libraries was much higher[146, 164, 168] although this does not seem to be universally the case[170]. This suggests that there is less value in sparsely searching the sequence space of variants with many changes compared with more thoroughly searching the sequence space of variants with few changes. This “small perturbation mutagenesis” strategy has also been used in [168].

The primary method of selecting substitutions to make at each selected position was from polymorphisms at that position in the R1a-B6 clonal lineage dataset.

Several of the mutations found in other branches of the clonal lineage may have been driver mutations which alone or in combination confer affinity or binding breadth improvements. Using only polymorphisms found in the R1a-B6 clonal lineage would have resulted in an insufficiently diverse combinatorial library. Most of the positions selected for mutation were either invariant within the clonal lineage sequence sample or had only one or two alternative amino acids at that position. Therefore, to increase the diversity of the library further, common residues found in unrelated alpaca VHH domains at the corresponding positions were also introduced. This data was collected from the web service Abysis[205], which provides the frequency distribution of amino acids at each position within antibodies. Residues listed on the Abysis database from Alpaca heavy chain domains at each selected residue with a frequency of >10% were recorded. At each mutated position, up to 5 desired mutations were listed, with preference given to the polymorphisms found in the R1a-B6 clonal lineage sample.

A handful of the desired mutations could not be introduced into the library due to the limitations of the synthesis method for the variable oligonucleotides. This method can change the distribution of nucleotides at a particular position in the oligonucleotide but cannot change the distribution of whole codons so mutations can only be introduced at high frequency if they differ from an unmodified codon by only one base. This limitation is explained in detail in table R1 4a,b,c,d,e. The full range of amino acid substitutions selected for introduction into the R1a-B6 variant library is shown in table R1 5a,b,c.

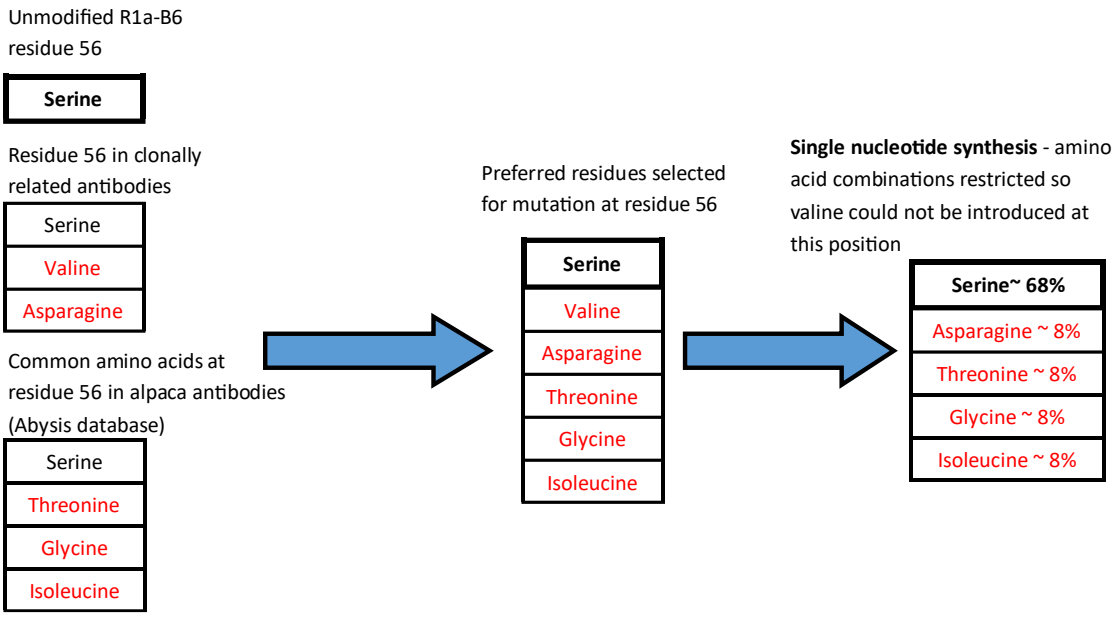
A)

TTT	Phe	TCT	Ser	TAT	Tyr	TGT	Cys
TTC		TCC		TAC		TGC	
TTA	Leu	TCA		TAA	STOP	TGA	STOP
TTG		TCG		TAG		TGG	Trp
CTT	Leu	CCT	Pro	CAT	His	CGT	Arg
CTC		CCC		CAC		CGC	
CTA		CCA		CAA	Gln	CGA	
CTG		CCG		CAG		CGG	
ATT	Ile	ACT	Thr	AAT	Asn	AGT	Ser
ATC		ACC		AAC		AGC	
ATA		ACA		AAA	Lys	AGA	Arg
ATG	Met	ACG		AAG		AGG	
GTT	Val	GCT	Ala	GAT	Asp	GGT	Gly
GTC		GCC		GAC		GGC	
GTA		GCA		GAA	Glu	GGA	
GTG		GCG		GAG		GGG	

B)

Unmodified residue	**Residues resulting from a single base substitution**	**Residues requiring 2 or more base substitutions**
Serine	Isoleucine	Phenylalanine
	Threonine	Leucine
	Asparagine	Methionine
	Cystine	Valine
	Arginine	Proline
	Glycine	Alanine
		Tyrosine
		Histidine
		Glutamine
		Lysine
		Aspartic acid
		Glutamic acid
		Tryptophan

C)



D)

A – 92%	**G – 76%**	**T – 100%**
G – 8%	T – 8%	
	C – 8%	
	A – 8%	

E)

Amino acid	Codon	Frequency (%)
Serine	**AGT**	69.9
Glycine	**GGT**	6.1
Isoleucine	**ATT**	7.4
Threonine	**ACT**	7.4
Asparagine	**AAT**	7.4
Valine	GTT	0.6
Alanine	GCT	0.6
Aspartic acid	GAT	0.6

Table R1 4a,b,c,d,e: Summary of library design considerations using residue 56 as an example The method used to mutate R1a-B6 genes in the combinatorial library was the introduction of base substitutions at low frequency at designated positions in and around the CDRs. Therefore, the occurrence of 2 different substitutions within the same codon would be rare. **A)** displays the possible codons and corresponding amino acids that can result from a single base substitution (yellow) in the codon AGT (green) which was selected to encode the R1a-B6 residue serine 56. Codons requiring 2 or more substitutions from AGT are coloured red. **B)** shows which residues can result from a single nucleotide substitution of AGT and which require 2 or 3 substitutions. Those on the former list can be incorporated into the designed selection of mutations introduced into R1a-B6 variants. **C)** The amino acids to introduce at each mutated residue position were selected primarily from clonally related antibodies. For example, at position 56, some related antibodies encoded valine or asparagine instead of serine. These residues were supplemented by common residues found at position 56 in other alpaca antibodies. Due to the constraints of the synthesis method, valine could not be introduced at position 56 at ~8% frequency. **D)** shows the codon encoding serine 56 and the frequency of mutant base

substitutions. In table **E)** the frequencies of all the possible amino acids encoded by the codon in **D)** are listed. The 4 selected amino acids, glycine, isoleucine threonine and asparagine (all in bold) are each introduced at near to 8% frequency but if two substitutions are introduced into the same codon, different amino acids are encoded, resulting in a much lower frequency of valine, alanine and aspartic acid substitutions.

CDR1 and surrounding residues

A)

Residue No.	26	27	28	29	30	31	32	33	34	35
WT residue	**G**	**S**	**F**	**F**	**S**	**R**	**Y**	**R**	**M**	**G**
Selected mutations (mutation frequency 8%)	E	R			T	T			I	A
		I			I	K			V	S
					R	I				
					N	S				

CDR2 and surrounding residues

B)

Residue No.	50	51	52	53	54	55	56	57	58	59	60	61	62	63
WT residue	S	I	A	Y	D	G	S	T	S	Y	A	D	P	V
Selected mutations (mutation frequency 8%)	G		T		G	D	N	P	N			G	S	A
	R		S		N	S	G	A	T			E		
	T		G				T	I						
							I							

CDR3

C)

Residue No.	93	94	95	96	97	98	99	101	102
WT residue	N	L	D	P	P	G	I	L	Y
Selected mutations (mutation frequency 8%)			N		S		N		S
			G		R				D
			A		A				

Table R1 5a,b,c: Mutations introduced into CDRs and surrounding residues with oligonucleotide base substitutions Residues in red are introduced into CDR encoding oligonucleotides by base substitution at a frequency of approximately 8% per mutation. For example, at residue 50 ~8% of the library would encode each of glycine, arginine and threonine and ~76% of the library would encode the unmutated residue, serine. **A)**, **B)** and **C)** show selected residue variation and frequency for CDR1, 2 and 3.

3.5.4. Testing the assembly and amplification process

Before yeast transformations were carried out, an experiment was conducted to ensure that the product of the soePCR and pull through PCR reaction was assembled in order, in frame and with the correct mutations inserted into the sequence. Full length mutagenised R1a-B6 genes (assembled and amplified as described in section 2.3.1 and 2.3.2) were recombined with a pNIBS-5 backbone by Gibson assembly (section 2.3.4) and transformed into *E.coli*[206]. As an additional control, the product of the same assembly reaction was also amplified

using a standard PCR kit (not epPCR) (section 2.3.3), recombined with pNIBS-5 backbone and transformed into *E.coli* too.

From each transformation reaction, 20 individual transformed colonies were picked and the plasmid segments containing the antibody genes were sequenced (section 2.1.2).

The sequencing showed correct mutations were being introduced at all of the targeted positions. Encouragingly, both the sequenced group with and without epPCR had average number of designed amino acid substitutions per gene (4.35 and 3.45 respectively) relatively close to the planned substitution frequency of 3.76. As expected, the average frequency of random substitution mutations in the epPCR group was higher than in the group without epPCR (0.45 and 0.20 respectively).

There was a high frequency of single nucleotide deletions found in both sequenced groups. These mutations result in a frameshift partway through the gene, resulting in a likely non-functional antibody and a loss of the C-terminal SV5-tag needed for yeast display normalisation. These frameshift mutations were found throughout the gene in no clear pattern. They occurred in regions where assembly oligonucleotides overlapped as well as non-overlap regions and regions derived from both variable and constant oligonucleotides. The polymerase used to assemble and amplify the non-epPCR sequences has been used in previous sequencing in this project without the same issue. The cause of these deletions is unknown and may be due to defects in the synthesised oligonucleotides.

3.5.5. Scaled-up library transformation into yeast

Despite the frameshifting issue, the sequencing experiments showed that the majority of the assembly product was correctly assembled, in frame and encoding R1a-B6 gene variants with the correct inserted mutations. A large-scale library transformation into yeast was carried out. The library transformation into yeast took advantage of the yeast homologous recombination DNA repair system. The amplified mutagenised VHH genes overlapped with pNIBS-5 backbone so if both

were transformed into yeast by heat shock, the two components would be recombined into a complete circular plasmid.

Yeast were transformed in 16 separate reactions with an additional control reaction. 0.75µg of pNIBS-5 backbone (see section 2.1.8) and 3µg of epPCR amplified mutant insert was added to each of the 16 transformation reactions. The control reaction contained vector only (0.75µg of pNIBS-5 backbone with no insert). By plating serial dilutions of pooled transformed yeast, the number of unique transformation events was estimated to be 1.4 $x10^6$ per reaction or 2.3 x 10^7 total (this number will be referred to as the unsorted library size). In comparison, the vector only control had 970, indicating that the >99.9% of the library transformants contained plasmids containing inserted DNA. The pooled library was grown in selective media to preclude growth of untransformed cells and stored in 20% glycerol at -80ºC.

3.5.6. Illumina sequencing and analysis of the unsorted library

The R1a-B6 variant library was characterised using Illumina sequencing. This technology allows the parallel sequencing of hundreds of thousands of individual DNA molecules in a heterogeneous sample of purified DNA, making it ideal for quantifying the frequency of each targeted mutation in the library as well as those inserted during epPCR amplification[207]. A sample of the library was grown in 6 ml SD/CAA medium and plasmid DNA was extracted from this liquid culture. The region of the pNIBS-5 plasmid including the VHH gene was amplified in a 2-step nested PCR reaction with the second set of primers overhanging to add adapter sequences to each end of the amplified product to facilitate Illumina sequencing. PCR product was purified and quantified. Sequencing was carried out as recommended by Illumina. The protocol for amplification and purification of VHH genes is described in more detail in section 2.5.

In addition to the heterogeneous VHH library yeast population, a single colony of EBY100 yeast encoding R1a-B6 on pNIBS-5 was grown up in liquid culture and the R1a-B6 gene was amplified and sequenced with the other samples. The sequence data from this sample served as a baseline for the level of sequence variation introduced by the yeast growth, PCR amplification and sequencing read

errors. By subtracting this baseline level of variation from library samples, the frequency of mutations can be estimated while minimising the systematic error from these sources of sequence noise. Section 2.5.2 describes how this baseline control was implemented in more detail.

Figure R1 7 summarises the Illumina sequence data analysis pathway used to estimate the frequencies of mutations and whole clones found in this library. All analysis was carried out in Geneious Prime (Build 2019-11-07). As shown in the flow chart, sequences are read in the forward direction from the start of the VHH gene and in the reverse direction starting at the end of the gene. The maximum read length was 300bp from the ends of the adapter sequence, but most reads terminated before the maximum length. Consequently, no single read could cover the entirety of the VHH gene (348bp) and forward and reverse reads had to be matched to each other and merged to create a full length VHH sequence. Forward and reverse read matching was carried out with the Geneious “merge reads” function which searches for reverse complementary sequences on forward and reverse reads. The maximum read overlap was 208 bases but given that most sequences terminated before the 300bp maximum, an overlap of 100bp was more typical. The major limitation of this algorithmic sequence merging process is the inevitability of mismatches in which the forward read on one sequence is matched with the reverse read corresponding to a different sequence[208]. Given the sequenced libraries are all composed of variants of the same VHH gene with strong sequence similarity, such hybrid sequences are likely to be common.

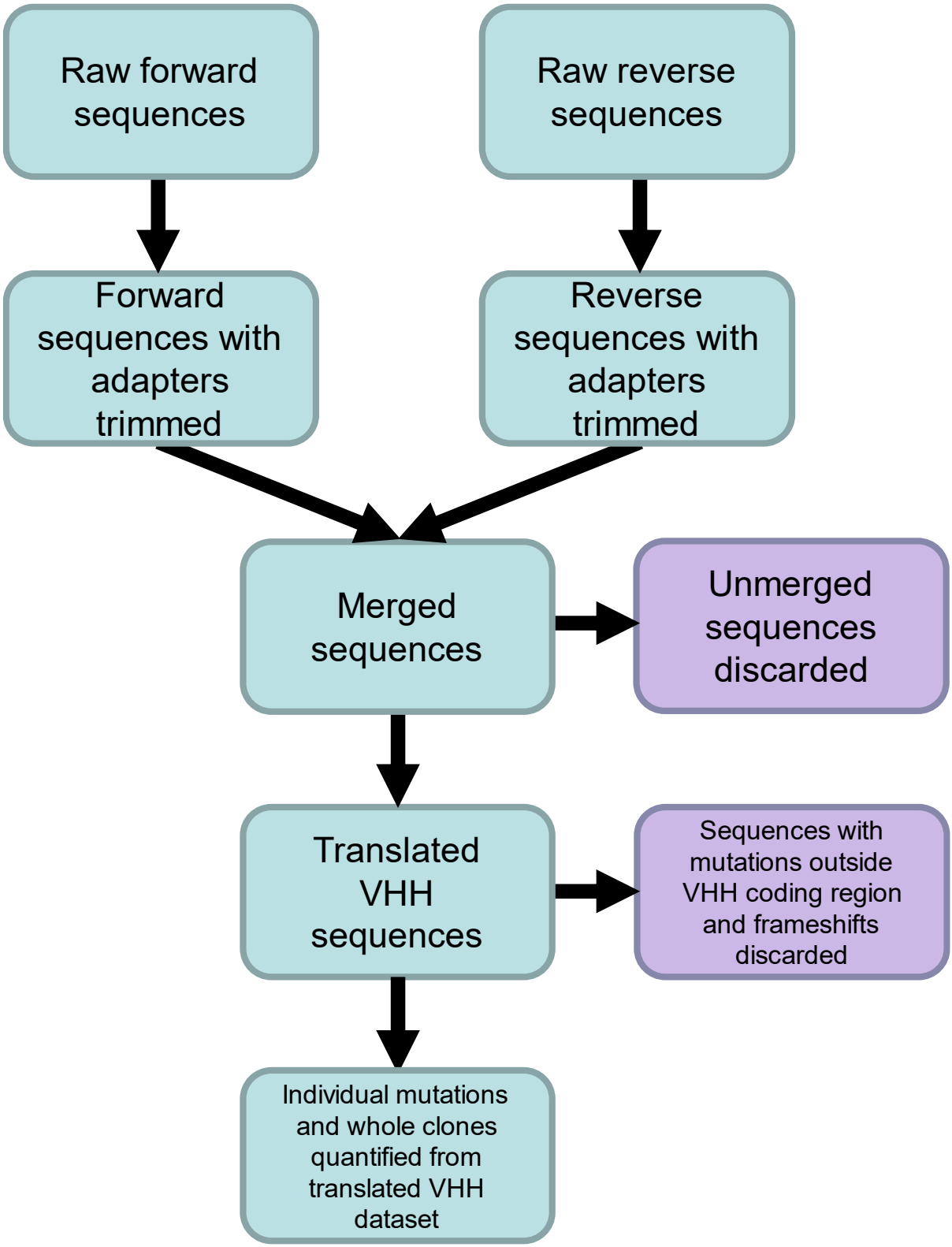


Figure R1 7: Outline of the processing of R1a-B6 variant library Illumina sequencing data Illumina sequencing was used to produce hundreds of thousands of reads in both directions for each of the R1a-B6 variant libraries. The maximum read length was not long enough to completely cover the VHH gene so forward and reverse reads were matched up computationally before being merged into a single continuous full-length sequence and translated. Sequences with mutations in the regions not coding for VHH were discarded, as were sequences with VHH

coding regions of the incorrect length. After this filtering, DNA sequences were translated into amino acid sequence. Frequencies of individual mutations and whole clones were calculated as a proportion of the number of translated sequences.

Hybrid sequences should be kept in mind whilst analysing Illumina sequencing data as they can introduce systematic bias into results. For example, the frequency of a rare individual clone is likely to be underestimated in a merged dataset as some of the forward and reverse sequences from this clone will have been erroneously combined with the wrong reverse or forward read. On the other hand, the frequency of a single mutation in a dataset (e.g. R31I) is unlikely to be affected by erroneous recombination so the changes in frequency of individual mutations can be treated with more confidence.

3.5.7. Analysing the diversity present within the unsorted library and estimation of the number of unique transformants

Due to the high frequency of deletions in trial library sequencing data, it was desirable to estimate the frequency in the full scale library of insertion and deletion mutations, most of which would result in a frameshift. To this end, the proportion of merged sequences with a length unequal to 392bp (the correct full length merged read length) was determined. To adjust for insertions and deletions introduced in the sequencing process, the proportion of incorrect length R1a-B6 control reads was subtracted from this value. Thereby, the proportion of VHH genes with insertion and/or deletion mutations in the mutagenised library was estimated to be 37.5%. It is likely that most of these mutations prevented HA binding and/or correct SV5-tag display and therefore they would have reduced library effective diversity as yeast expressing these VHH would not be selected in cell sorting experiments.

To estimate the frequency of each of the 47 targeted mutations in the mutagenised library (prior to sorting), the merged, translated and filtered (figure R1 7) Illumina sequencing files from the mutagenised library and the R1-B6 control were both queried to find the percentage of sequences with the mutant amino acid at the correct position (e.g., to find the frequency of R31I, frequency of I at position 31 was queried, irrespective of surrounding sequence). The

percentages from the R1a-B6 control file were subtracted from the mutation percentages from the mutagenised library file. For example, R31I was found at 8.3% frequency in the mutagenised library sequencing file and 0.6% in the R1a-B6 control so the estimated frequency in the yeast mutagenised library was 7.7%. Figure R1 8a compares the frequencies of mutations in the library design with the calculated frequencies in the Illumina data. Most mutations are within ±20% of their designed frequency and for none of the mutations does the discrepancy between designed and observed frequency exceed 2-fold. By summing the frequencies of this set of 47 mutations, the mean number of targeted mutations per molecule can be estimated for the designed library and the observed sequencing data. There were an estimated 3.43 targeted mutations per molecule in the library design and in close agreement, 3.38 were found per molecule in the sequenced sample.

Next, the frequency of all amino acid substitution mutations (not just targeted substitutions) in the mutagenised library was estimated. A random sample of 10,000 sequences was taken from the set of merged, translated and filtered sequences of the mutagenised library and of the R1a-B6 control. For each, the sample of sequences were sorted and tabulated (table R1 6) by number of substitutions (i.e. number of sequences with 0 substitutions, 1 substitution, etc) and from this table, a mean number of substitutions was calculated in each sample. By subtracting the mean substitutions in the R1a-B6 control file from the mean substitutions in the mutagenised library file, the mean number of amino acid substitutions per VHH gene in the mutagenised library was estimated to be 4.26.

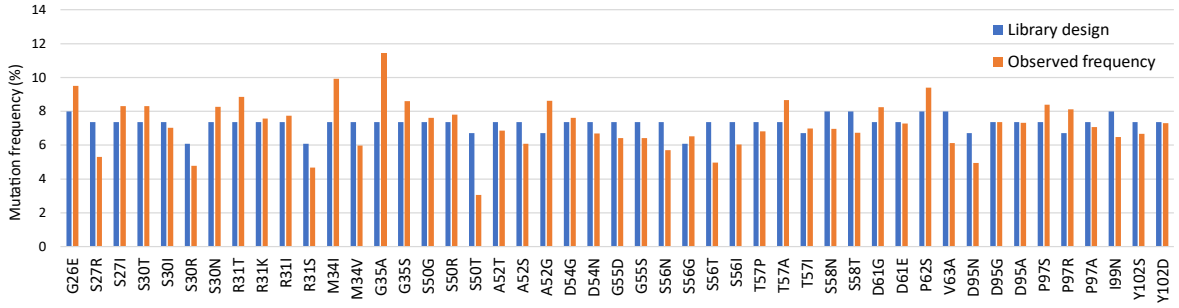
It was observed that the distribution of number of substitutions per gene closely resembles the Poisson distribution with the same mean. Therefore, the distribution of targeted substitutions per molecule and total substitutions per molecule could be modelled using the Poisson distribution, adjusting for mutations introduced in the sequencing process. Figure R1 8b shows the estimated number of sequences per 10,000 sequences containing 0, 1, 2… etc amino acid substitutions per molecule. The Poisson means used were 3.43, 3.38 and 4.26 corresponding to, respectively, number of targeted mutations (library

design), number of targeted mutations (Illumina sequencing data) and all amino acid substitutions (Illumina sequencing data).

The number of permutations of n targeted mutations can be estimated with ${}_nC_{47}$. Given that each individual mutation is independent and present at near uniform frequency, each of these permutations will also have an approximately equal frequency. Therefore, proportion of the permutations with n targeted mutations found in the library in in frame sequences (referred to as L) can be estimated from the number of in frame library variants with n targeted mutations (referred to as V) and the calculated number of permutations, again using the Poisson distribution. The following formula was used $L_n = 1 - Pois(X=0)$ where $\lambda = V/{}_nC_{47}$. Using this approximation, it can be estimated that >99.9% of 0, 1, 2, 3 and 4 targeted mutation permutations are present in the library in in frame sequences, 69% of 5 mutation combinations and 9% of 6 mutation combinations (see figure R1 8c). Furthermore, the number of targeted mutation combinations present in the library in frame as a whole can therefore be estimated from $L_0 \times {}_0C_{47} + L_1 \times {}_1C_{47} + L_2 \times {}_2C_{47} \ldots$, resulting in a figure of 3.1×10^6.

This estimate is a lower bound and likely a substantial underestimate of the true effective diversity of the unsorted library as it only accounts for targeted mutations and not for mutations introduced by epPCR.

A)



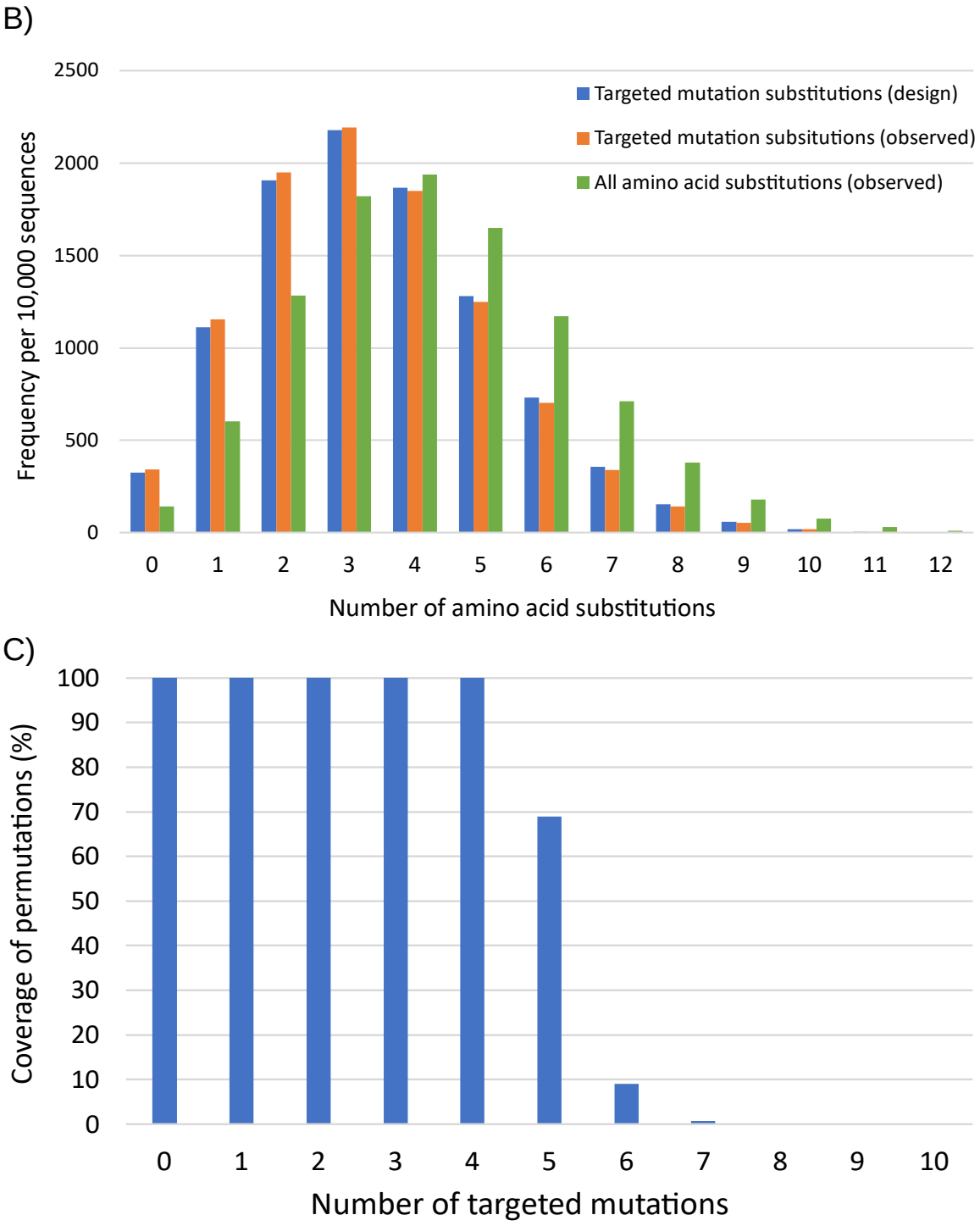


Figure R1 8a,b,c: Statistical analysis of the diversity withing the R1a-B6 variant library A) comparison of designed (blue) and observed (orange) frequency of each of the 47 targeted mutations in the R1a-B6 variant library. **B)** Estimated distribution of number of mutations per VHH molecule in 10,000 mutagenised sequences modelled using the Poisson distribution. According to the library design, a mean of 3.43 targeted mutations per VHH molecule. 3.38 targeted mutations per VHH were observed. Including all amino acid substitutions, 4.26 mutations were observed per VHH gene. **C)** Estimated percentage targeted mutation permutation coverage by in frame library sequences.

Number of mutations	Number of sequences	
	R1a-B6 control	Unsorted library
0	7478	64
1	1843	352
2	476	921
3	142	1618
4	33	1987
5	18	1841
6	7	1329
7	2	902
8	0	429
9	0	206
10	0	78
11	0	27
12	0	11
Mean mutations per sequence	0.35	4.61

Table R1 6: Distribution of number of amino acid substitutions in 10,000 randomly selected sequences from the R1a-B6 control and the unsorted library Illumina sequencing files

3.5.8. Flow cytometric characterisation of R1a-B6 variant library

To further verify that the R1a-B6 variant library had been transformed into yeast and was displaying correctly, a sample of this library was induced to express VHH and analysed by flow cytometry. The staining protocol and range of recombinant HA used was identical to section 3.2 except 50nM HA was used, not 12.5nM. Yeast expressing unmodified R1a-B6 and yeast expressing the non-functional R1a-B6 variant F28A (positive and negative controls respectively) were stained and analysed by flow cytometry alongside the R1a-B6 variant library. The resultant plots are shown in figure R1 9. As observed previously, R1a-B6 measurably bound H1, H5 and H9, resulting in almost all the VHH expressing population in quadrant 2. Meanwhile, for R1a-B6 stained with H2 and H3 and for all the F28A mutant displaying populations, yeast expressing VHH were overwhelmingly found in quadrant 3 due to weak or absent binding. Yeast expressing the R1a-B6 variant library stained with H1, H5 and H9 showed a diffuse cloud of events spanning quadrants 2 and 3, demonstrating that a range of H1, H5 and H9 binding affinities are found in the library. This is encouraging

as it demonstrates that the genetic changes in the library have led to a range of phenotypic changes to the antibodies' binding properties. By contrast, the H2 and H3 stained library resembled H2 and H3 stained R1a-B6 and F28A expressing yeast. There were few events in quadrant 2, indicating that detectable H2 and H3 binding was rare and would require enrichment to be observed.

In comparison to the R1a-B6 and F28A encoding yeast, the library sample contained a far higher proportion of yeast in quadrant 4 (neither expressing VHH nor binding HA). ~60% of library populations compared with ~20% of control populations were detected in this quadrant. It seems likely that this reflects the large number of frameshifted mutants within the library. Frameshifted VHH genes will produce products which do not bind HA (therefore do not stain with AF647) nor possess the C-terminal SV5-tag required for AF488 staining.

Overall, flow cytometric analysis demonstrated that over a third of the stained R1a-B6 variant library yeast cells were expressing VHH at a high level with a wide range of H1, H5 and H9 binding properties, providing a rich starting point for selection of more broadly cross-reactive R1a-B6 variants.

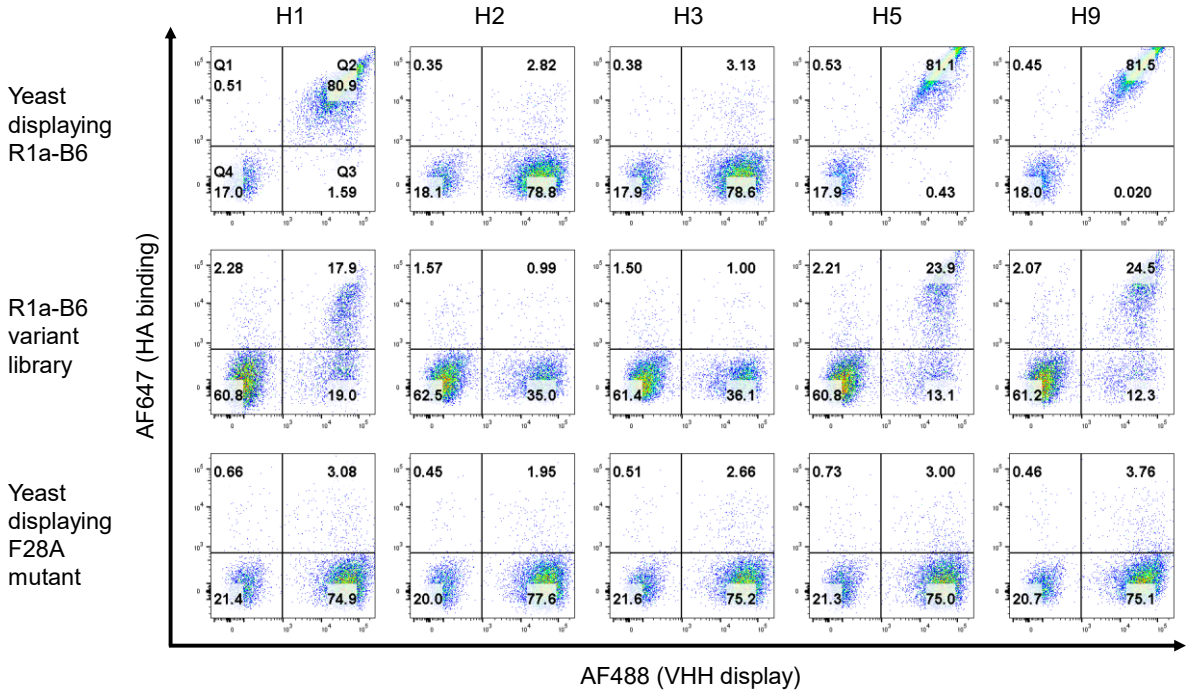


Figure R1 9: Flow cytometry dot plots of the R1a-B6 variant library stained with HA The HA binding properties of the R1a-B6 variant library (middle) was compared to monoclonal population of yeast displaying R1a-B6 (top) and F28A R1a-B6 mutant (bottom). To highlight the differences in staining patterns, the dot plots are split into 4 quadrants (Q1-4) and each quadrant is labelled with the percentage of events inside that quadrant.

4. Results chapter 2 – Yeast display library sorting and isolation of R1a-B6 variants

4.1. Chapter introduction

With a library expressing over 3 million unique, in-frame variants of R1a-B6, the next major challenge of the project was to separate out the small number of variants, potentially just a handful, with a significantly broader range of HA antigen reactivity. The challenges of sorting the highest affinity variants of an antibody in a large library using yeast display are addressed in [139, 197, 201, 209] and sorting protocols used in this project are based on these works.

A library of antigen variants displayed on the surface of yeast is a mixture containing a vast number of subpopulations expressing antibodies with a range of affinities. Through several rounds of sorting, the aim was to enrich the few subpopulations with the highest affinity for the antigen used for selection and to deplete subpopulations with lower affinity to extinction. If this is achieved, random sampling of the final sorted library is likely to yield high affinity variants. A complication is that even a genetically homogeneous population of yeast will display a variable amount of antibody. Furthermore, both staining of an individual yeast cell and the fluorescence intensity measurement of that cell by flow cytometry is subject to statistical noise. Therefore, even yeast cells expressing identical amounts of identical antibody will be measured to possess a range of fluorescence intensity values when stained with antigen. To mitigate the issue of noise in yeast display staining levels, yeast were stained with one fluorophore corresponding to the level of antigen binding and one corresponding to the level of antibody display. Thereby, variable antibody display levels can be compensated for in the gating strategy, maximising retention of higher affinity sub populations and depletion of lower affinity sub populations[201, 209]. This concept is illustrated in figure R2 1.

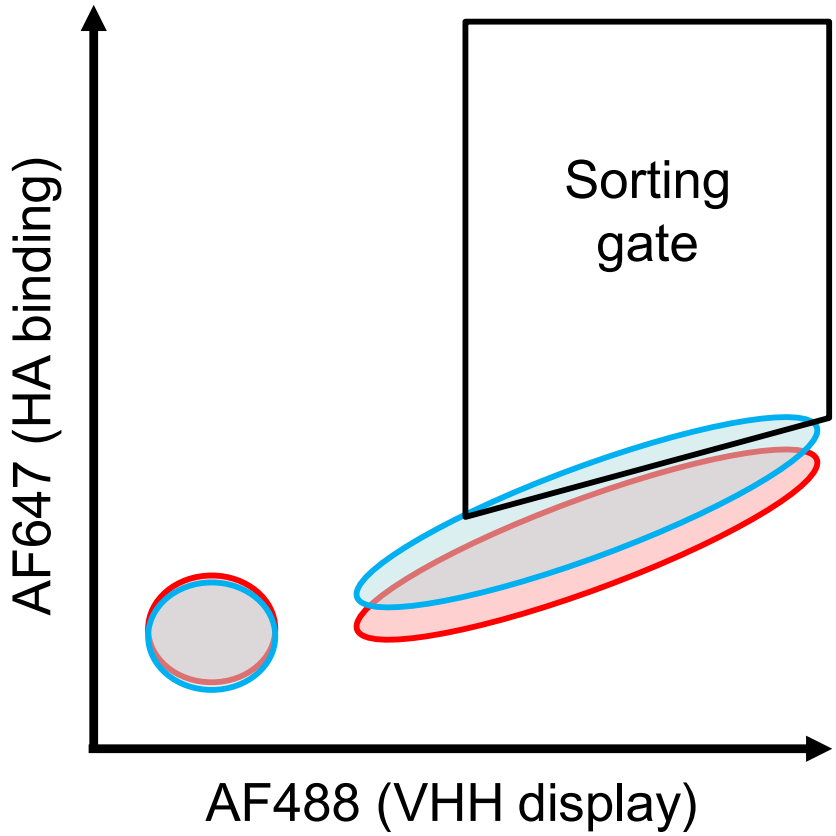


Figure R2 1: Cartoon depicting the use of a diagonal sorting gate to maximally enrich a higher affinity subpopulation FACS library sorting experiments aim to enrich higher affinity subpopulations and deplete lower affinity subpopulations. In this figure, the gating strategy aims to enrich the higher affinity blue population from the lower affinity red population. Both display VHH to the same extent so have highly similar AF488 staining distributions but yeast cells in the blue population displaying VHH bind more HA so the blue population stains with greater AF647 fluorescence. Due to statistical noise in staining levels and fluorescence measurement, red and blue populations appear as overlapping clouds of dots on the plot (represented as ovals in the figure). A sorting gate with a diagonal lower boundary normalises for the reduced AF647 staining in cells expressing less VHH, maximising retention and enrichment factor of blue cells over red cells.

Another consideration in yeast display library sorting is to ensure, especially in the first sorting round, that rare, improved subpopulations do not go extinct by chance. To achieve this, the number of cells to be sorted in the first-round library is substantially in excess of the size of the unsorted library, ideally by approximately an order of magnitude. If this is the case, the average number of copies of a unique transformant will be 10 and an overwhelming majority of the unique transformants (97%) will be present at 5 or more copies. Therefore, failure to express, under-staining, etc is unlikely to, by chance, result in the loss of a high affinity variant[201]. In subsequent rounds of sorting, it was assumed that high

affinity subpopulations had already been substantially enriched so fewer cells were stained and sorted to save on reagents and sorting time.

In addition, the first round of each sorting campaign used a sorting mode called “yield” which attempts to maximise the retention of cells in the gated region at the expense of inclusion of a small number of extra cells outside the gate being retained. After first round sorts, the assumption is that higher affinity variants have already been enriched many-fold so in subsequent rounds, the sorting mode is switched to “purity”. “Purity” mode prioritises exclusion of cells outside the gate at the expense of loss of a small proportion of yield of cells inside the gate.

For the same reasons, sorting gates were drawn to be relatively permissive in the first sorting rounds and increasingly stringent in subsequent sorting rounds. For example, in H9 round 1 sort, 1.0% of stained yeast cells were retained while sorts H9 R2 and H9 R3b retained 0.5% and 0.1% respectively. A permissive gate improves the chances that a rare, improved clone will be retained in the initial sort while in later steps, stringent gates enrich higher affinity clones to a greater degree[201].

Figure R2 2a shows a flow chart of the sorting rounds intended to select for high affinity H2 and H9 binding variants and the resultant sub-libraries and table R2 1 shows the numbers of cells and sorting conditions of each sorting experiment. Sub-libraries were named after the sorting round which produced them. For example, H9 round 1 sorting experiment produced H9 round 1 sorted library (H9R1).

4.2. Staining protocol design

The staining setup used for library sorting is illustrated in figure R2 2b and the protocols used for each sort are described in detail in section 2.4 (table MET 7b staining protocols 2a, 2b, 2c). Briefly, yeast cells were incubated with mouse α-SV5-tag antibodies, followed by AF488 goat α-mouse antibodies. These two steps served to stain the library with AF488 according to VHH surface display levels. The final staining step used a mixture of biotinylated recombinant HA

(A/Canada/720/2005 for H2 selections, A/Hong Kong/1073/99 for H9R1,2,3a selection and A/Guinea fowl/Hong Kong/WF10/99 for the H9R3b selection) and an excess of streptavidin AF647. These two reagents were premixed to bind to each other to produce HA-fluorophore complexes which were in turn intended to bind to surface expressed VHH.

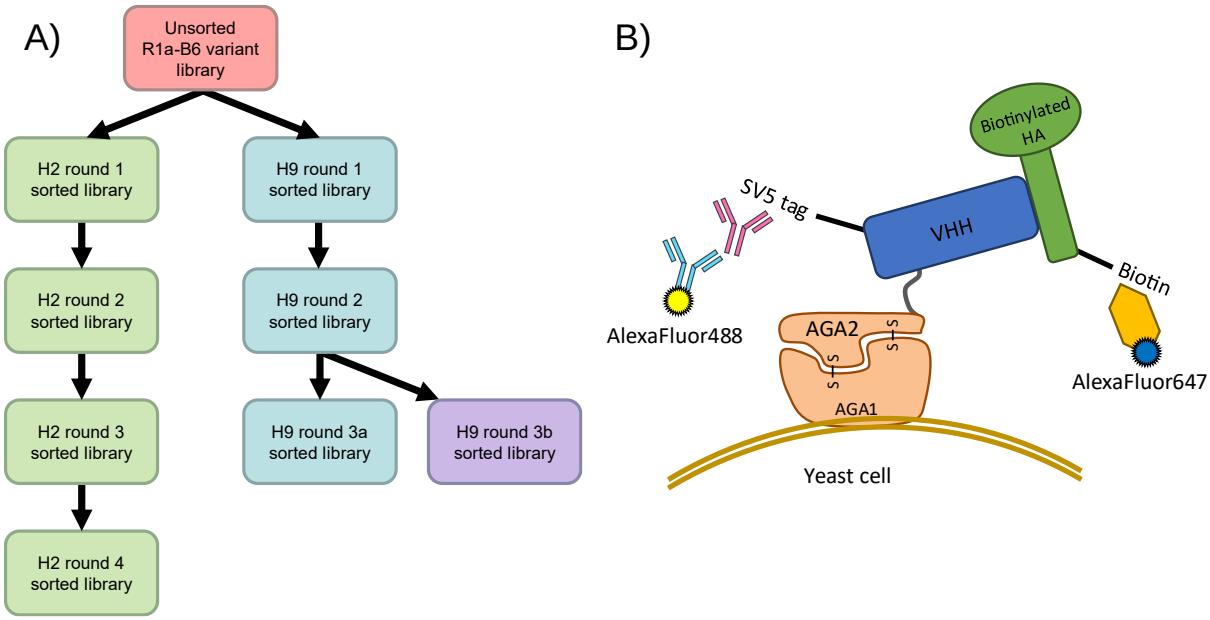


Figure R2 2a,b: sorting of the R1a-B6 variant library for binding to H2 and H9 HA A) flow chart depicting the sub libraries sorted from the unsorted library (red) in campaigns to isolate R1a-B6 variants with superior binding properties to H2 (green) and H9 (original 3 sorted libraries H9R1, H9R2 and H9R3a in blue, H9R3b library sorted by amended method in purple) **B)** Cartoon depicting expressed VHH on a yeast cell surface stained for cell sorting. Protocols used to stain yeast for sorting are described in detail in section 2.4. Yeast sorted in all H2 sorting steps were stained with staining protocol 2a. Yeast for sorts H9R1,2 and 3a were stained by protocol 2b and for sort H9R3b yeast were stained with protocol 2c.

Sort	Source population	Cells sorted	Percent of cells retained	Purity or yield mode?
H2R1	Unsorted library	1×10^8	1.5	Yield
H2R2	H2R1	1×10^7	0.9	Purity
H2R3	H2R2	1×10^7	0.1	Purity
H2R4	H2R3	1×10^7	0.3	Purity
H9R1	Unsorted library	2×10^8	1.0	Yield
H9R2	H9R1	1×10^7	0.5	Purity
H9R3a	H9R2	1×10^7	0.2	Purity
H9R3b	H9R2	1×10^7	0.1	Purity

Table R2 1: Library sorts carried out on R1a-B6 variant libraries

Recombinant HA (H2-A/Canada/720/2005 and H9-A/Hong Kong/1073/1999) was biotinylated using a commercial kit according to manufacturer's instructions and biotinylation was verified using the HABA assay. H9-A/Guinea fowl/Hong Kong/WF10/1999 recombinant antigen was sold in biotinylated form so did not need to be biotinylated like the other HA antigens.

For the staining protocol used in results chapter 1 (table MET 7b, staining protocol 1), cells were stained with HA in the first of 3 staining steps. By contrast, in this chapter in library sorting experiments (staining protocols 2a, 2b, 2c), yeast cells were stained with HA in the third of 3 staining steps. It was anticipated that a shorter time gap between washing excess HA off stained cells and sorting would increase sensitivity as HA would have less time to dissociate from surface expressed antibody.

The staining protocol was designed with the idea in mind that yeast could be simultaneously co-stained with two different biotinylated HAs pre-incubated with two different streptavidin fused fluorophores. This would have resulted in a 3-colour stain. One fluorophore for VHH display, and one for binding to each of the two HAs. This would aid in selection for broadly cross-reactive variants.

4.3. Data collection on R1a-B6 variant libraries

Sorted sub-libraries were sequenced with Illumina sequencing by the same protocols as the unsorted library and the R1a-B6 control (see section 2.5). Likewise, the methods of data processing and analysis were identical too. As the unsorted library was parental to all of the sorted libraries, the extent to which mutations had been enriched or depleted in successive sorting steps could be calculated by dividing the adjusted frequency of a mutation in a sorted library by the adjusted frequency in the unsorted library. In the context of affinity maturation selection, the enrichment of a mutation in successive rounds of selection is suggestive that it improves affinity for antigen. Conversely, depletion of a mutation from sorted libraries implies that it is detrimental to affinity for antigen.

Supplementing data collection by Illumina sequencing was Sanger sequencing of isolated colonies randomly picked from yeast VHH libraries. The VHH genes from 32 colonies from each of the sorted libraries and the unsorted library were sequenced (as described in section 2.1.3). Sequence data by this method was much faster to obtain so was used for preliminary analysis.

4.4. Selection for H2 binding

The first set of library sorts used H2 recombinant antigen (A/Canada/720/2005) for staining. Binding of R1a-B6 or the unsorted library to this antigen had not been detected using flow cytometry (see figure R1 9) but R1a-B6 has shown binding to antigen of the highly similar H2 isolate A/Singapore/1/1957 in previous ELISA data and bivalent R1a-B6 has demonstrated neutralisation of a live influenza virus with an A/Singapore/1/1957 haemagglutinin protein[103]. Therefore, there was reason to believe that there was a low affinity interaction between R1a-B6 and the H2 HA stem region.

The very weak existing interaction between R1a-B6 and H2 antigen (below the limit of flow cytometric detection) made selection of improved variants very challenging. A yeast displayed R1a-B6 variant may need to interact with H2 with substantially higher affinity than the unmodified VHH to be stained detectably more with AF647 on the flow cytometer.

Nonetheless, the selection was attempted. To maximise the chance of identifying improved H2 binding variants, a high concentration (100nM) of biotinylated H2 antigen was used in the staining protocol and a large excess (10-fold) of streptavidin AF647 was used to ensure saturation of the biotin bound to HA.

The first sort used 10^8 cells from the unsorted library, an approximately 5-fold excess of unsorted library size. The subsequent 3 sorts used 10^7 cells each but the same staining protocol with reagent volumes scaled down. Dot plots of each sort and the gates used for them are shown in figure R2 3 and details of the sorts are displayed in table R2 1. The first 3 sorts showed little or no improvement in AF647 staining. However, the 4^{th} round of sorting revealed a significant

subpopulation staining with AF647 correlating with AF488 fluorescence, as would be expected for a VHH variant binding to H2 antigen.

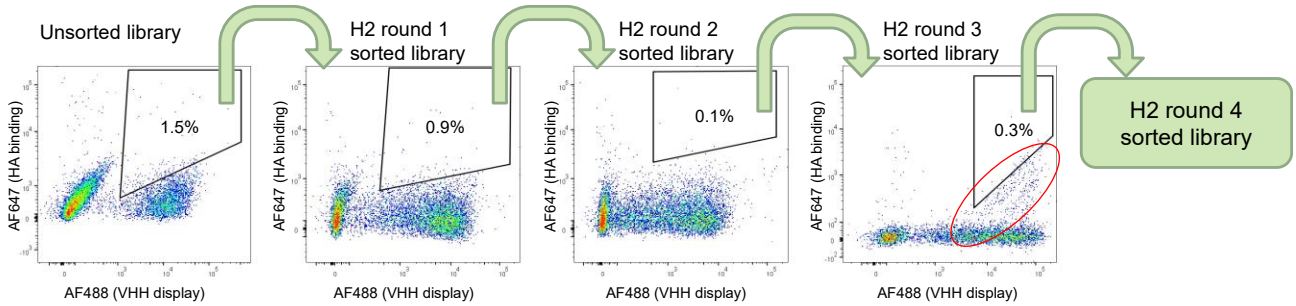


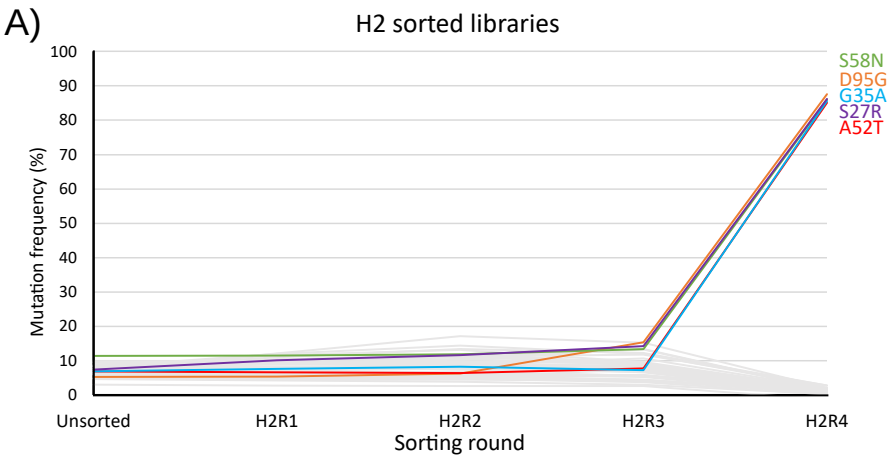
Figure R2 3: H2 library sorting gates and dot plots Dot plots displaying a sample of 10,000 events from each population stained to be sorted. Gates used to sort the cells are also shown with the percentage of events inside the gates labelled. Highlighted in the red circle in the round 4 sort is an AF647 staining subpopulation of the round 3 sorted library.

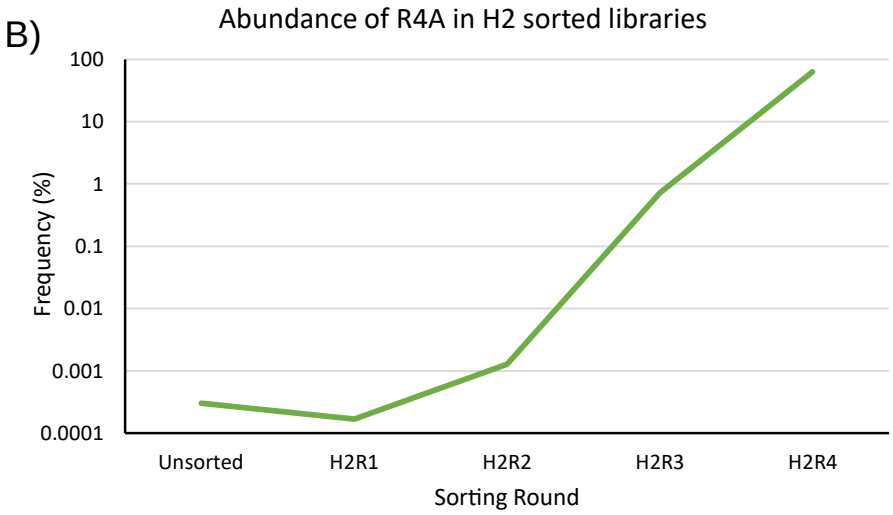
Preliminary sequence data revealed that a majority (21/32 of Sanger sequenced clones) of the H2R4 sorted library consisted of a single clone, henceforth referred to as R4A (H2R4 clone A). R4A contained 5 of the deliberately inserted amino acid substitutions (S27R, G35A, A52T, S58N, D95G) as well as the mutation S70Y which was likely introduced by epPCR. This clone was assumed to be responsible for the AF647 staining subpopulation within H2R3 which was seen during the H2 R4 sort.

In agreement with the Sanger sequencing data, Illumina sequencing revealed that clone R4A was enriched by an estimated 350,000 fold between H2R1 and H2R4 sorted libraries to a final frequency of 62%.

Estimation of the frequencies of the set of 47 targeted mutations in the sorted libraries revealed little enrichment or depletion in the first 3 sorting steps. The maximum enrichment factor found in the first 3 rounds of selection was 2.9 and the minimum was 0.61, suggesting that none of the mutations were individually highly advantageous or disadvantageous in allowing a yeast cell to be sorted into the next round of selection. Conversely, between H2R3 and H2R4, the 5 targeted mutations (as well as S70Y) found in R4A jumped to a frequency of around 85%, while all other targeted mutations were heavily depleted. This reflects the jump of R4A from a minority sub population in H2R3 to the majority of the library in H2R4. Frequencies of targeted mutations in the unsorted library and libraries H2R1-4 are displayed in figure R2 4a, the frequency of clone R4A is tracked

throughout the H2 sorting rounds in figure R2 4b and the sequence of clone R4A is compared to R1a-B6 in the regions in and around CDRS 1, 2 and 3 in figure R2 4c.





C) CDR1 and surrounding residues

	26	27	28	29	30	31	32	33	34	35
R1a-B6	G	S	F	F	S	R	Y	R	M	G
R4a	G	R	F	F	S	R	Y	R	M	A

CDR2 and surrounding residues

	50	51	52	53	54	55	56	57	58	59	60	61	62	63
R1a-B6	S	I	A	Y	D	G	S	T	S	Y	A	D	P	V
R4a	S	I	T	Y	D	G	S	T	N	Y	A	D	P	V

CDR3

	93	94	95	96	97	98	99	101	102
R1a-B6	N	L	D	P	P	G	I	L	Y
R4a	N	L	G	P	P	G	I	L	Y

Figure R2 4a,b,c: Enrichment of individual mutations and clone R4A in H2 sorted libraries Illumina sequencing was performed on the unsorted library and each H2 sorted library to provide information on the mutations and whole clones enriched in the sorting process. Sequences with mutations outside the VHH encoding gene and sequences in which the VHH gene was the wrong length due to insertions or deletions were filtered out and the frequencies of mutations were calculated as a proportion of the number of remaining sequences. **A)** depicts the frequency of each of the set of 47 introduced mutations into R1a-B6 in each of the Illumina sequencing datasets. None of the individual mutations appear to be greatly enriched or depleted from the

unsorted library to H2 R3. However, between H2R3 and H2R4 the mutations S58N, D95G, G35A, S27R and A52T are each enriched from a frequency of under 20% to approximately 85%. The frequency traces of these five mutations are coloured green, orange, blue, purple and red respectively and the frequency trace of every other mutation is coloured grey. These mutations are enriched due to the enrichment of clone R4A which jumps from less than 1% frequency to over 60% frequency in between the round 3 and round 4 sorted libraries **B)**. The data also shows that this clone was enriched in sorting rounds 2 and 3 as well. Each of the enriched individual mutations are found in the sequence of clone R4A. **C)** shows the sequence of R4A at CDRs 1,2 and 3 aligned to R1a-B6. Mutations are highlighted according to the colour scheme in **A)**.

4.4.1. Investigation of clone R4A and the H2R4 sorted library

Following sorting experiments, the dominant clone of the H2R4 library, R4A, was isolated and characterised alongside R1a-B6 using yeast display and flow cytometry. Yeast populations induced to display VHH were stained with a range of HA antigens (H1, H2, H3, H5, H9) at 50nM concentration using flow cytometry staining protocol 1 (section 2.4.). It was noted that R4A displaying yeast stained with AF647 (used to stain for HA binding) to a similar extent with each of the HA antigens tested. This result could be explained if R4A had a very similar affinity for each of the HA antigens. Alternatively, AF647 staining observed could be independent of the presence of HA antigen.

Further experiments were carried out including the H2R4 sorted library as well as additional controls and the results are summarised in figure R2 5a. Staining was carried out with 50nM antigen (where applicable) according to staining protocol 1. As expected, R1a-B6 showed strong binding to H1 antigen and no detectable binding to H2 antigen. When the HA was removed entirely from the staining protocol, R1a-B6 expressing yeast did not stain with AF647. Both R4A and a majority subpopulation of the H2R4 sorted library stained with an approximately equal intensity with H1, H2 and no HA present, indicating that clone R4A stains with AF647 independently of HA.

Another control used H1 HA and substituted the streptavidin AF647 reagent with streptavidin AF405 (used at a 1:250 dilution) which should bind to biotin as streptavidin AF647 does and therefore be functionally equivalent in a staining protocol. AF405 is a dimmer fluorophore so the signal is weaker. Nonetheless, yeast cells expressing R1a-B6 clearly stain with the substituted fluorophore. On

the other hand, neither the H2R4 sorted library, nor the R4A expressing yeast showed detectable AF405 staining with this protocol.

The hypothesis arising from this set of results was that clone R4A and therefore a majority of the H2R4 sorted library did not bind detectably to H1 but instead bound directly to the fluorophore streptavidin AF647. This hypothesis is depicted in figure R2 5b. Furthermore, based on the lack of detectable AF405 staining when library H2R4 was stained with the fluorophore substituted, no sizable subpopulation of library H2R4 retains high affinity H1 binding.

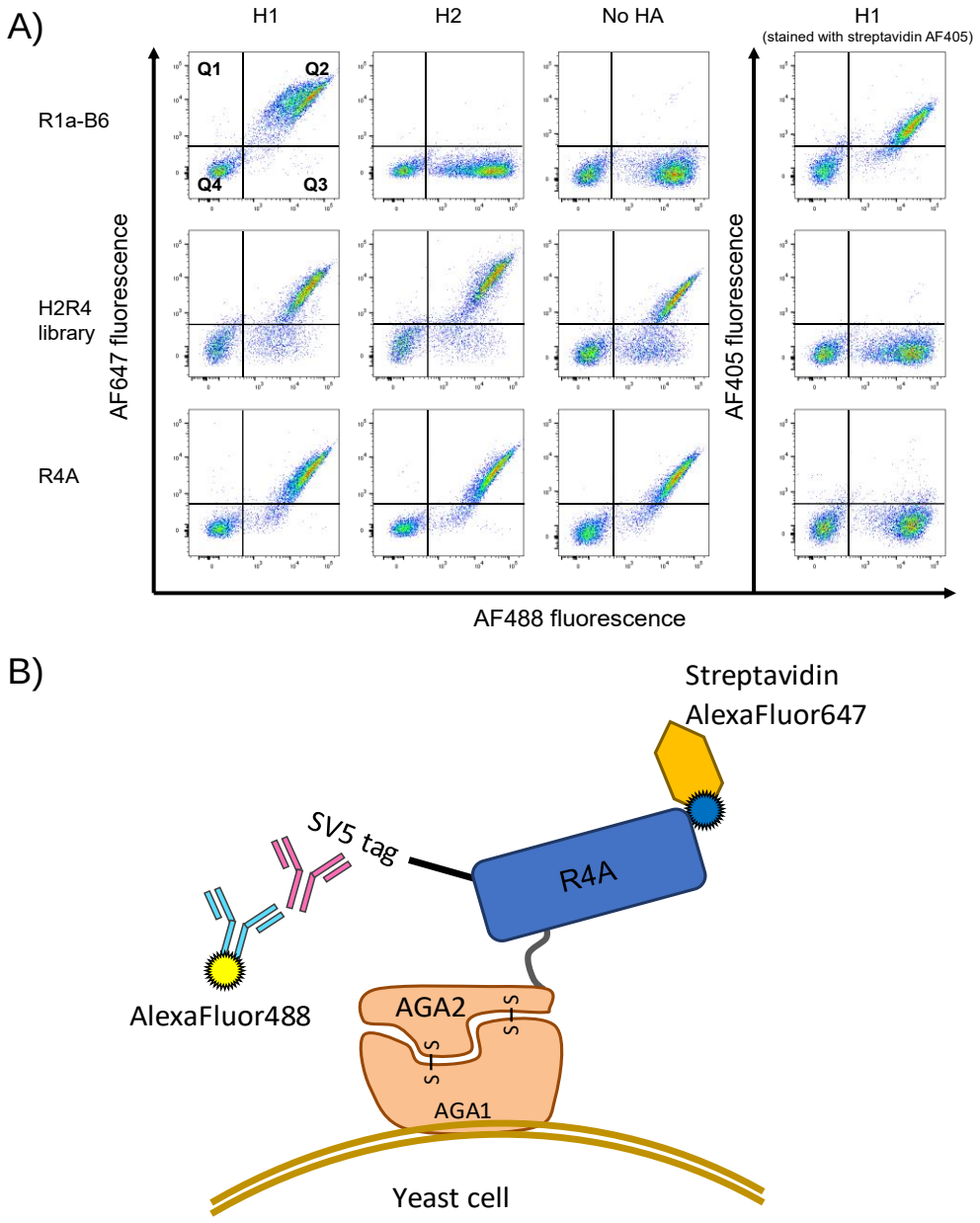


Figure R2 5a,b: Investigation of H2R4 library and clone R4A A) Flow cytometry data showed that a majority subpopulation of the H2R4 library and the clone that made up the majority of the H2R4 library, R4A, was staining with AF647 independently of HA (indicated by many events in Q2). AF647 staining is usually HA dependent. In addition, R4A and this subpopulation of the H2R4 library did not stain when streptavidin AF647 was substituted out for streptavidin AF405 in staining protocol 1 (1:250 diluted) (these dot plots show many events in Q3 but few in Q2). Meanwhile, the AF647 staining of yeast displaying R1a-B6 is dependent on HA of subtypes H1, H5 or H9. Furthermore, R1a-B6 displaying yeast will stain when the streptavidin AF647 is substituted for streptavidin AF405. HA in this experiment was at a 50nM concentration. The conclusion reached from this data was that variant R4A bound directly to streptavidin AF647 and any binding to H1 was below the limit of detection. **B)** depicts a schematic of the hypothesised interaction at the stained yeast cell surface when R4A is displayed. As with other VHH, AF488 staining is facilitated by SV5-tag display. Meanwhile R4A binds directly to streptavidin AF647 not to HA.

It is likely that during the H2 library sorting steps, the significant excess of streptavidin AF647 used to stain the libraries meant that the streptavidin AF647 binding VHH variant, R4A, was enriched to a greater extent than any variants capable of binding H2 antigen. Unfortunately, this possibility had not been considered during experimental design. Subsequent sorting experiments used substantially less streptavidin AF647 in the staining protocol to reduce the extent to which streptavidin AF647 binding clones could be enriched and to instead favour selection of high affinity HA binding clones.

4.5. Selection for H9 binding

Running in parallel to H2 sorting experiments, the unsorted library was selected for H9 antigen binding in 3 rounds of selection, referred to as H9R1, H9R2 and H9R3a (figure R2 2a), resulting in the sorted library H9R3a. Figure R2 6a shows flow cytometry dot plots of the 3 rounds of sorting and the gates used in the sorts. Details of the sorts can be seen in table R2 1. Libraries were stained according to flow cytometry staining protocol 2b (section 2.4) which differs from the staining protocol used for H2 library sorts in two ways. The antigen used was biotinylated recombinant H9 antigen (A/Hong Kong/1073/1999) and the concentration of the antigen was halved to 50nM (with a corresponding halving of the streptavidin AF647 reagent to maintain the same excess ratio). Due to the timing of these sorts, the presence of streptavidin AF647 binding variants such as R4A in the

library was not yet realised so the protocol was not amended at this stage to mitigate selection of these variants.

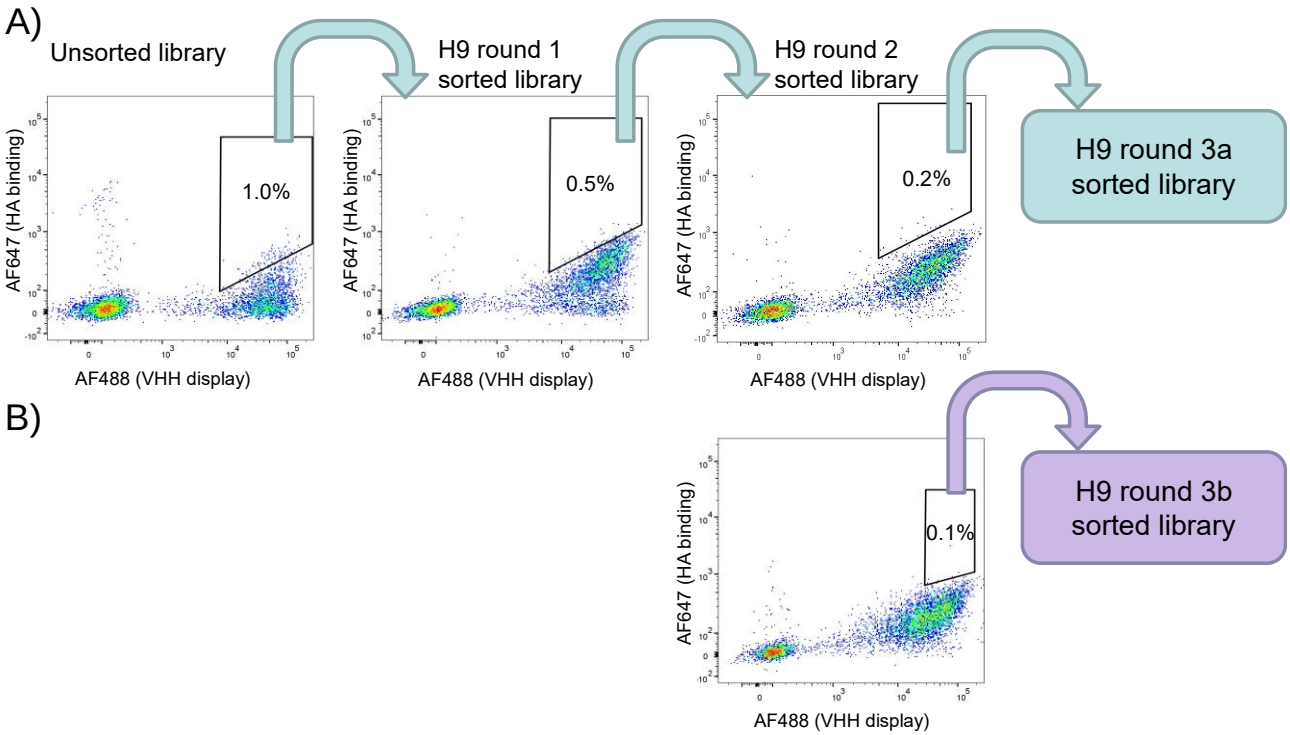


Figure R2 6a,b: H9 library sorting gates and dot plots Dot plots displaying a sample of 10,000 events from each population stained to be sorted. Gates used to sort the cells are also shown with the percentage of events inside the gates labelled. **A)** H9 library sorts H9R1-3a (staining protocol 2b) **B)** H9 library sort 3b (amended staining protocol 2c).

Sort H9R1 used 2 x 10^8 cells (an approximate 10-fold excess over the unsorted library size) and 1.0% of the sorted cells were retained. Sorts H9R2 and H9R3a both used 10^7 cells and 0.5% and 0.2% of cells were retained respectively. The dot plots in figure R2 6a show a progressive increase in the AF647 staining (HA binding) of the AF488 positive (VHH displaying) yeast populations as well as a reduction to almost zero of the population showing high AF488 fluorescence but no AF647 fluorescence (yeast displaying VHH but not binding HA).

Initial Sanger sequencing of library H9R3a revealed that of 32 sequences, 17 were identical to at least one other sequence in the sample (by comparison, all unsorted library Sanger sequences were unique). There were four groups of identical sequences in the sample, referred to henceforth as G1, G2, G3 and G4, appearing in the sample 6, 6, 3 and 2 times respectively. The presence of repeated identical sequences in a sample of H9R3a indicated that these clones

were highly enriched in the library sorting process so the yeast colonies encoding these variants were isolated.

4.5.1. Investigation of library H9R3a and clones selected in sorts H9R1-R3a

As with H2 sorting selections, H9 sorted libraries were compared to yeast expressing R1a-B6 using flow cytometry. A minority population of yeast in H9R3a were found to stain with streptavidin AF647 independently of HA indicating the presence of streptavidin AF647 binding clones. Figure R2 7a shows overlayed dot plots of yeast expressing R1a-B6 and library H9R3a both stained with protocol 1 (section 2.4.) with HA omitted from the protocol. In addition, it was also noted that the H9 R3a library showed noticeably less AF488 fluorescence than its parent library H9R2 (figure R2 7b).

To investigate the phenomenon of low AF488 staining, the Sanger sequence data from H9R3a was re-examined and it was noted that the sequence encoding the SV5-tag sequence was mutated in 27 of 32 sequences including the sequences of G1-4 (by comparison, no sign of SV5-tag mutation was found in the Sanger sequences from the unsorted library). 3 new clones from the H9R3a library sanger sequenced sample were isolated and characterised by flow cytometry alongside yeast expressing R1a-B6 and G1-4. Clone 14 (C14) had an unaltered SV5-tag sequence, while C23 and C27 both had different mutations in their SV5 tag sequence. All these clones were stained according to flow cytometry staining protocol 1 (section 2.4.) with 12.5nM HA (H1, H5 and H9 subtypes) as well as without HA.

A strong correlation was revealed between mutations in the SV5-tag sequence and a reduced AF488 MFI (figure R2 7c). Yeast expressing R1a-B6 and C14, both with unmodified SV5-tags had very similar AF488 MFI values. Meanwhile, G1-4, C23 and C27, all with mutated SV5-tags, showed AF488 MFI values reduced by 40-60% compared with R1a-B6. It was hypothesised that mutated SV5-tags would bind to mouse α-SV5-tag antibodies with a lower affinity and would therefore be more likely to dissociate before measurement, reducing the AF488 fluorescence of the cell without reducing the level of displayed VHH

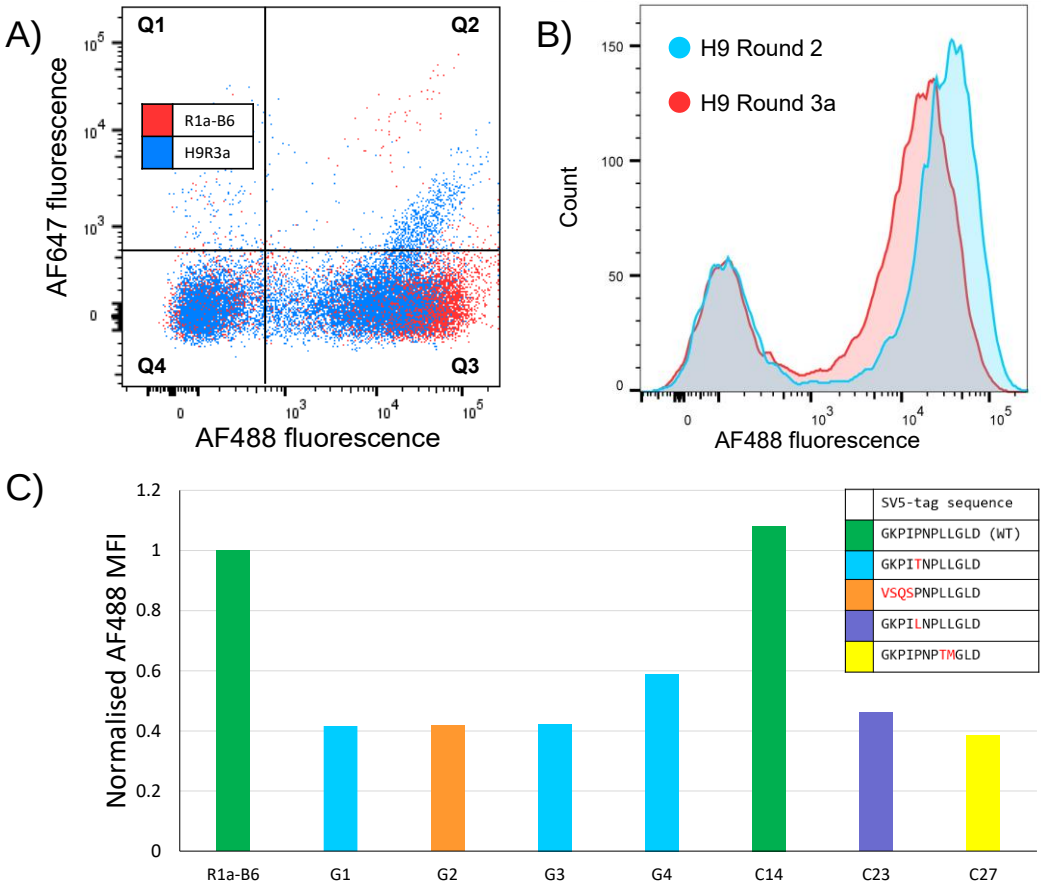
(illustrated in figure R2 7d). The fact that a majority of library H9R3a was composed of SV5-tag mutant clones implied that these clones were heavily selected for in the sorting process. This phenomenon can be explained by consideration of the gating strategy used, especially the diagonal lower boundary of the gate. Figure R2 7e shows the H9R3a sorting dot plot with the gate used to sort the cells annotated. One factor which favours a yeast cell to land inside the sorting gate is display of a VHH with high affinity for HA, which tends to increase the quantity of HA binding to the cell and therefore the AF647 staining (blue arrow in figure). Reduction in the level of AF488 staining without reduction in the level of actual VHH display and AF647 staining provides an alternative route for a yeast cell to land inside the gate (red arrow). Therefore, with this gating strategy, yeast with mutated SV5-tags, improved H9 affinity or a combination of these two factors were selected for in the sorting steps.

A similar unintended artefact of yeast display library selection was observed in [147]. In this study a Myc epitope tag was used for display normalisation not an SV5-tag. Clones isolated from yeast display affinity maturation selections included clones with mutations in the Myc-tag sequence resulting in a decrease in Myc-tag dependent staining and a selective advantage in the sorting process.

Of the 7 R1a-B6 variants investigated in this experiment, only one, C14, showed HA independent staining and this clone presumably comprised part of the HA independent streptavidin AF647 staining population observed in the whole library sample.

Despite the unintended selection for SV5-tag mutants and streptavidin AF647 binding variants, the general trend in the group of clones tested was an increase in H9 binding compared to R1a-B6 (figure R2 7f). An at least 2-fold increase in MFI was seen in all variants except G4 and C14. C27 was the most improved H9 binder at 3.5-fold increase relative to R1a-B6. The same 5 antibodies showed more modest improvements to H5 antigen binding and similar or slightly decreased binding to H1 antigen compared to R1a-B6.

Based on this preliminary binding data, the cell sorting process seemed to have been partially successful. 5 different antibody clones were isolated with improved H9 binding as measured by flow cytometry and all 5 retained substantial binding to H1 and H5 subtype antigens. Furthermore, the fact that all these 5 clones possessed SV5-tag mutations, which are assumed to be rare in the unsorted library, implied that the full potential of the R1a-B6 variant library for H9 binding improvement had not been reached. Only a small fraction of the VHH library, those possessing SV5-tag mutations, had been effectively searched in these sorting experiments. Therefore, the decision was made to resume sorting from the H9R2 sorted library with an amended protocol to minimise the inclusion of unwanted clones such as streptavidin AF647 binding VHH and clones with SV5 tag mutations.



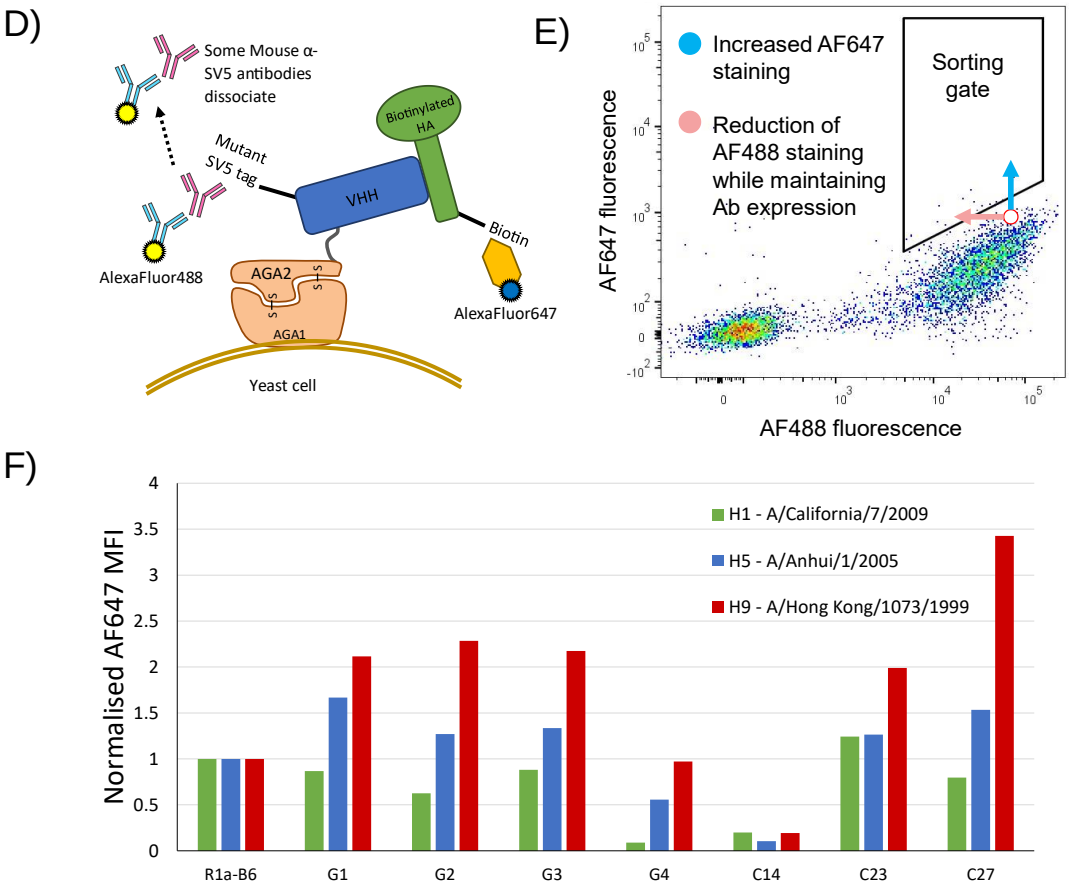


Figure R2 7a,b,c,d,e,f: Investigation of library H9R3a and variants isolated from library H9R3a A) dot plot of library H9R3a and R1a-B6 displaying yeast stained according to flow cytometry staining protocol 1 with no HA. As with library H2R4, library H9R3a contains an HA-independent staining population hypothesised to bind directly to streptavidin AF647. This population is, however, a minority of the library. **B)** histogram of AF488 staining of library H9R2 compared to H9R3a stained according to protocol 1. It was noted that between library H9R2 and library H9R3a, there was a major drop in AF488 staining. **C)** To investigate this phenomenon further, 7 R1a-B6 variants from library H9R3a were isolated and the AF488 MFI was measured after staining. 6 of the 7 variants showed reduced AF488 staining compared to yeast displaying R1a-B6. In addition, it was found that these 6 variants all had mutations in their SV5-tag sequence while the only variant with normal AF488 staining had an intact SV5-tag sequence. **D)** It was hypothesised that SV5-tag mutations reduced affinity of the tag for α-SV5 antibody, resulting in a reduction in AF488 staining. **E)** The diagonal gating strategy used in sorts H9R1, H9R2 and H9R3a favoured both increased AF647 staining (blue arrow) and decreased AF488 staining (red arrow). Increased AF647 staining could be achieved by increased HA binding or streptavidin AF647 binding. Meanwhile, SV5-tag mutations decreased AF488 staining. **F)** The same 7 R1a-B6 variants were stained with 12.5nM HA of H1, H5 and H9 subtypes and the AF647 fluorescence was measured. Most of the variants showed increased H9 binding compared to R1a-B6. Except for G4 and C14, H5 binding was maintained or slightly improved and H1 binding maintained or slightly worsened.

4.5.2. H9 sort 3b with amended protocol

In the previous sorting experiments, two factors were identified as probable sources of unintended sorting artifacts. Firstly, an excessive concentration of streptavidin AF647 favours the selection of VHH which can bind to this

fluorophore. A reduction in the concentration of this reagent in the staining protocol would reduce the AF647 fluorescence of yeast expressing fluorophore binding VHH. It was found that a fivefold reduction in streptavidin AF647 concentration had minimal impact on staining intensity of variant libraries or yeast displaying R1a-B6 and this reduced concentration was used in the new sort (referred to as sort H9R3b). Secondly, the shape of the gate used in sorting round H9R3a (figure R2 6a) favoured the selection of SV5-tag mutants as illustrated in figure R2 7e. Figure R2 6b shows a dot plot of the stained library H9R2 during the H9R3b sort with the amended sorting gate. The leftmost gate boundary is shifted to the right (moved to a greater value of AF488 fluorescence) and the lower boundary has a shallower gradient. Both these changes to the gate shape disfavour yeast with reduced AF488 fluorescence.

In addition to these protocol changes, due to reagent availability, recombinant H9 HA from a different influenza isolate (A/Guinea fowl/Hong Kong/WF10/1999) was used instead of the H9 HA (A/Hong Kong/1073/1999) used in sorting rounds H9R1-3a. However, the sequence difference is minimal and distant from the stem region epitope of R1a-B6[102]. Full details of the protocol used for this sort are found in section 2.4 (staining protocol 2c).

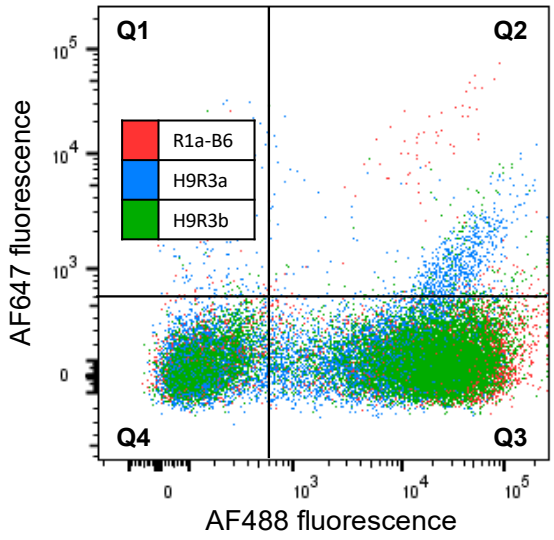
After sorting, the library was compared to yeast expressing R1a-B6, the unsorted library and previous H9 sorted libraries. Figure R2 8a shows an overlay of yeast expressing R1a-B6, libraries H9R3a and H9R3b, all stained according to protocol 1 with no HA. HA independent AF647 staining was clearly seen in the H9R3a library but was negligible in the H9R3b sorted library, validating the effectiveness of the reduction in fluorophore concentration in the staining protocol 2c.

AF488 MFI was also measured and shows that the H9R3b library maintains the level of AF488 staining of its parent library H9R2 (figure R2 8b). By contrast, library H9R3a drops noticeably in AF488 MFI compared to H9R2 indicating that the problem of SV5-tag mutant enrichment was substantially mitigated by improvements in protocol. Supporting this conclusion, Sanger sequencing of a sample of 29 members of library H9R3b showed only 5 mutated SV5-tag sequences.

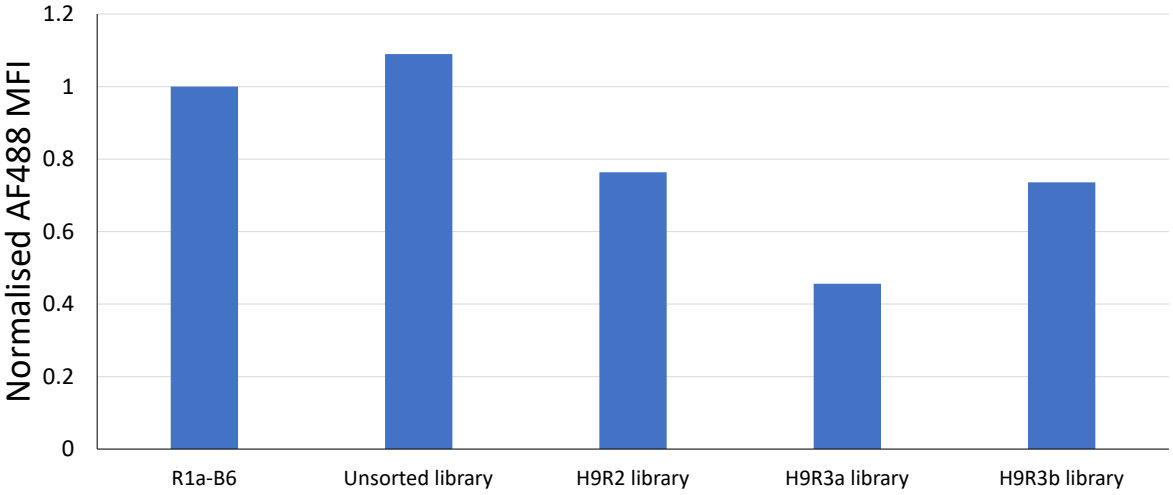
Finally, library H9R3b outperforms library H9R3a in binding to H1, H5 and H9 antigen as measured by flow cytometry (figure R2 8c). These results suggest that the improvements in sorting protocol not only reduced the occurrence of the unintended sorting artifacts identified in previous sorts but also increased the frequency and perhaps quality of improved H9 binding VHH variants.

Sanger sequencing data from library H9R3b returned no repeated sequences so R1a-B6 variant clones were selected at random. The 4 clones selected for further characterisation were named WC4, WC11, WC21 and WC25.

A)



B)



C)

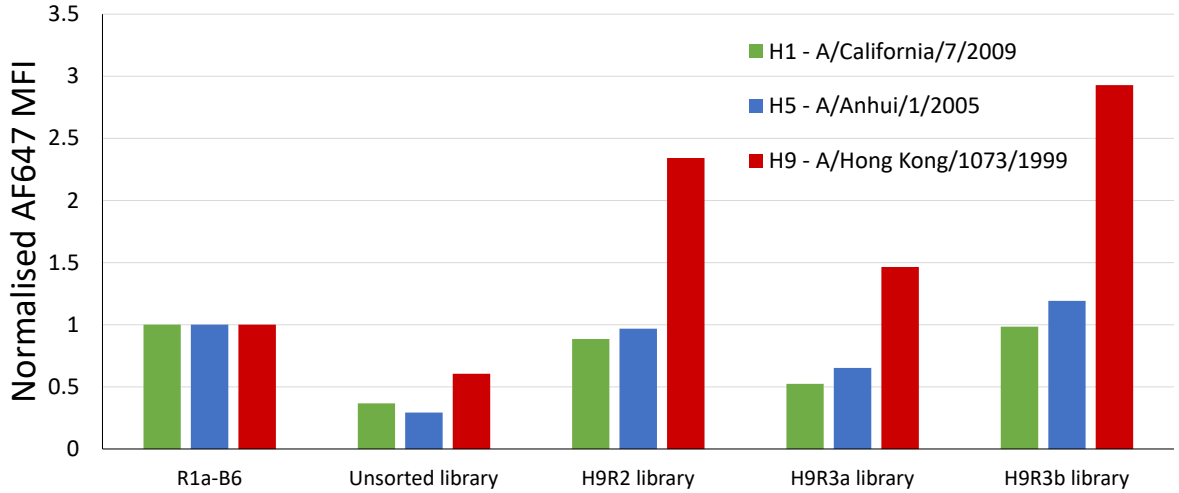


Figure R2 8a,b,c: Reduction of unintended sorting artifacts from H9R3b library with an amended staining and sorting protocol A) Overlayed dot plots of R1a-B6 displaying yeast (red), H9R3a (blue) and H9R3b (green) all stained according to protocol 1 without HA. The significant subpopulation of H9R3a which stains with AF647 independent of HA is not found or is overwhelmingly reduced in H9R3b. **B)** AF488 MFI is compared across H9 sorted libraries stained with protocol 1. H9R3a has a substantially reduced AF488 MFI compared to H9R2 while the AF488 MFI reduction is avoided in H9R3b which has a AF488 MFI nearly identical to that of H9R2. **C)** Comparison of ability of H9 sorted libraries to bind H1, H5 and H9 HA. Compared to the H9R2 sorted library, the H9R3a sorted library shows reduced binding to all 3 subtypes, likely reflecting enrichment of SV5-tag mutant and AF647 binding variants. Meanwhile, the H9R3b library shows increased binding to all 3 subtypes, reflecting an improved protocol which disfavours mutant SV5-tags and AF647 binding and selects for H9 binding improvements. To reduce the impact of the frameshift mutants found in the unsorted library, yeast cells with little or no AF488 staining (under a value of 500) were omitted from the datasets used to calculate MFI for figures **B)** and **C)**. In this experiment H9R1 library staining failed so this library was also omitted from results.

4.5.3. Illumina sequence analysis of H9 sorted libraries

As with the H2 sorted libraries, Illumina sequencing was used to track the frequencies of mutations and whole clones through the process of cell sorting using H9 antigen. Figure R2 9a,b,c shows the frequency changes through successive sorting steps of intentionally introduced mutations from the unsorted library to library H9R3a. The frequency changes of clones G1, G3, C23, C27 (the clones from library H9R3a which were successfully expressed and characterised in results chapter 3 (section 5)) and the sequences of these clones in the CDRs compared to R1a-B6 are also included in this figure. Figure R2 10a,b,c shows similar plots but using data from library H9R3b instead of H9R3a. Like in figure R2 9a,b,c, the VHH variants which were successfully expressed and characterised from library H9R3b, WC4, WC11, WC21 and WC25 are tracked in

frequency throughout the sorting process and their CDR sequences are compared to R1a-B6.

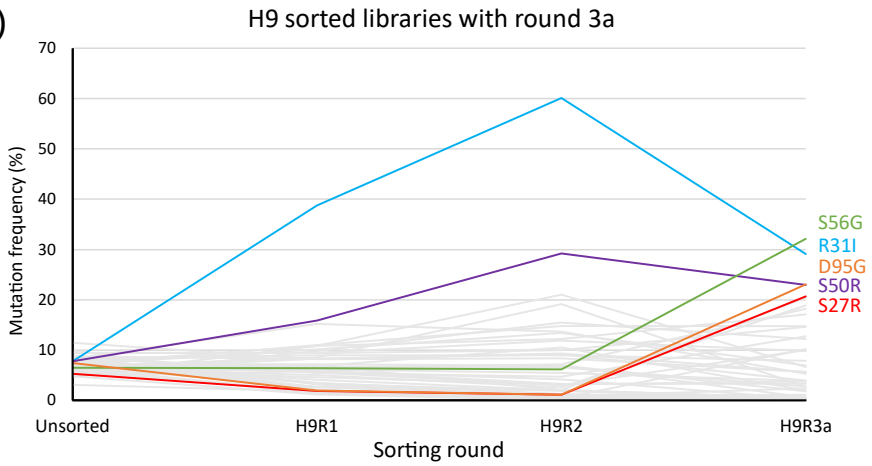
Unlike with the H2 library selections, the first two H9 antigen sorts (shared between the H9R3a and H9R3b branch of H9 sorted libraries) show clear evidence of enrichment and depletion of individual mutations. In library H9R2, mutation R31I is enriched 8-fold to a frequency of 60%, meanwhile, S30R is depleted almost to extinction by an estimated 150-fold. The final round of selection for the H9R3a branch (original sorting protocol) leads to the top 5 most frequent mutations being (in descending order) S56G, R31I, D95G, S50R and S27R. Interestingly, D95G and S27R were depleted in the rounds 1 and 2 sorts before enrichment in round 3. One possible explanation for this is an epistatic interaction in which these mutations may be detrimental on their own but advantageous to selection in particular combinations[210].

The most frequent mutations in library H9R3b (amended sorting protocol) are overlapping but not identical to those in H9R3a – R31I, R31S, S50R, D95N, P62S. Unlike in library H9R3a all 5 of these mutations were, to a greater or lesser extent, already enriched suggesting that each of these mutations were, on average, neutral or beneficial to antigen binding in many different mutation combinations.

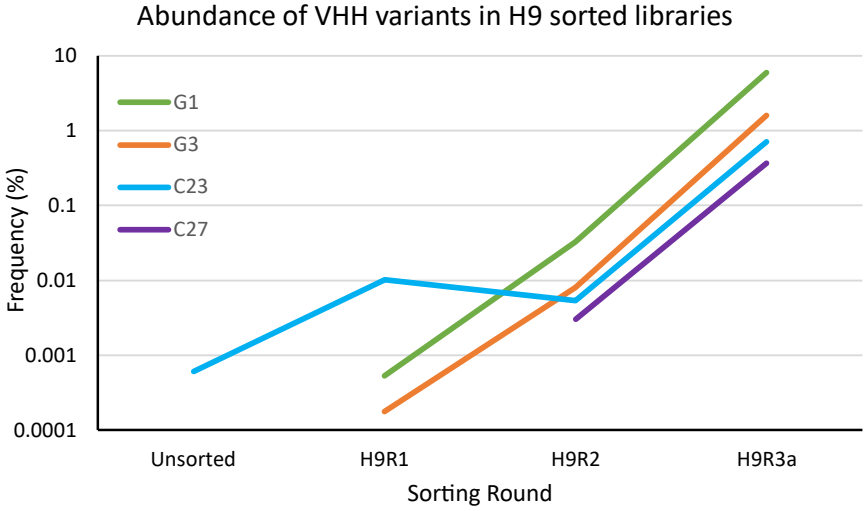
All 4 antibody clones from library H9R3a (G1, G3, C23 and C27) which were eventually expressed and characterised, show progressive enrichment in the selection process (figure R2 9b). Only C23 is seen to decrease in frequency between consecutive libraries. All 4 clones are enriched over 100-fold between H9R2 and H9R3a libraries, probably in part because of their SV5-tag mutations.

Likewise, the 4 clones expressed from library H9R3b (WC4, WC11, WC21 and WC25) can all be detected in the Illumina dataset from H9R1 onwards and 4 of them are progressively enriched (figure R2 10b). None of these clones are present at a detectable frequency in the unsorted library. In all cases, the final enrichment between libraries H9R2 and H9R3b is smaller than 100-fold. This may reflect a greater variety of improved clones being enriched in contrast to the small number of SV5-tag mutant clones enriched in sort H9R3a.

A)



B)



C)

CDR1 and surrounding residues

	26	27	28	29	30	31	32	33	34	35
R1a-B6	G	S	F	F	S	R	Y	R	M	G
G1	G	S	F	F	S	I	Y	R	M	G
G3	G	S	F	F	S	I	Y	R	M	G
C23	E	S	F	F	S	R	Y	R	M	G
C27	E	S	F	F	S	T	Y	R	M	G

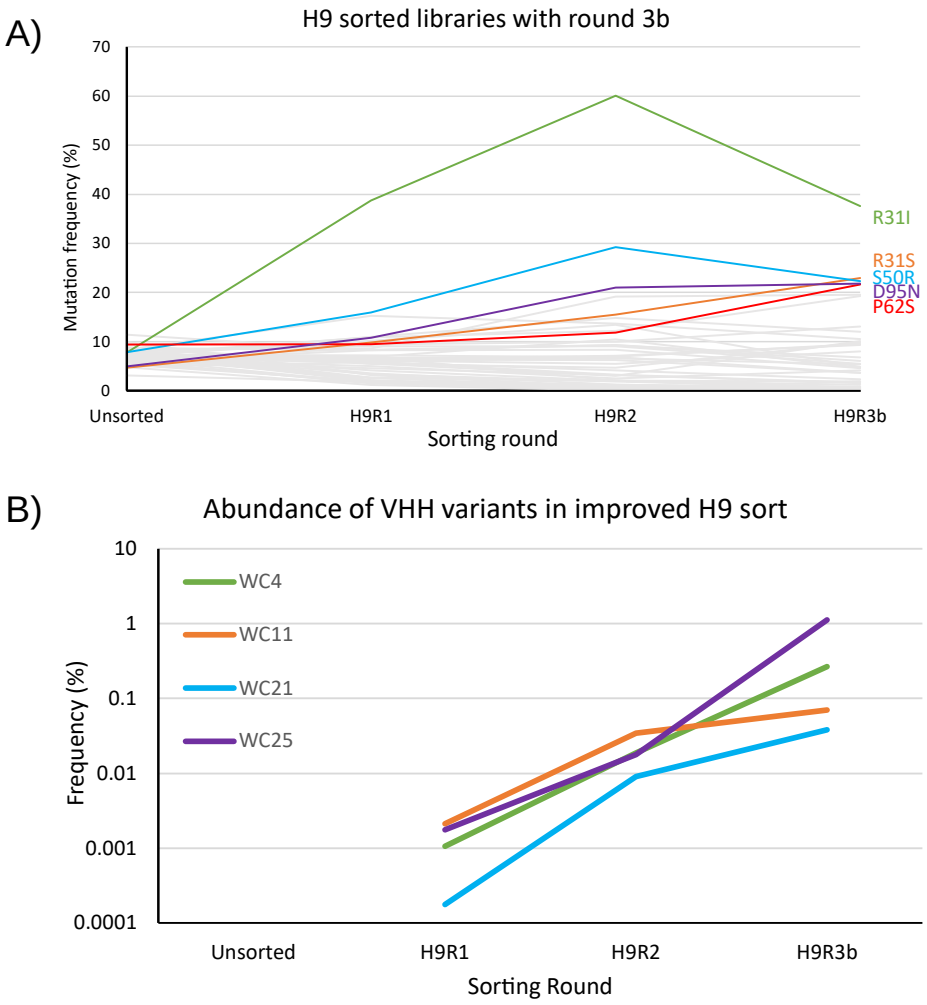
CDR2 and surrounding residues

	50	51	52	53	54	55	56	57	58	59	60	61	62	63
R1a-B6	S	I	A	Y	D	G	S	T	S	Y	A	D	P	V
G1	S	I	A	Y	D	G	G	T	N	Y	A	D	P	V
G3	S	I	T	Y	D	G	S	T	N	Y	A	D	P	V
C23	S	I	A	Y	D	G	I	T	S	Y	A	D	P	V
C27	S	I	A	Y	D	G	S	T	S	Y	A	D	P	V

CDR3

	93	94	95	96	97	98	99	101	102
R1a-B6	N	L	D	P	P	G	I	L	Y
G1	N	L	D	P	P	G	I	L	Y
G3	N	L	D	P	A	G	I	L	Y
C23	N	L	A	P	P	G	I	L	Y
C27	N	L	N	P	P	G	I	L	Y

Figure R2 9a,b,c: Enrichment of individual mutations and variants of interest in H9 sorted libraries H9R1, 2 and 3a A) As with figure R2 4 the frequency of each of the 47 intentionally introduced mutations is plotted against successive sorting rounds. In this figure enrichment is monitored through the H9 sorted libraries (H9R3a branch). The lines corresponding to the first, second, third, fourth and fifth most frequent mutations in library H9R3a (S56G, R31I, D95G, S50R, S27R) are coloured green, blue, orange, purple and red respectively. **B)** The frequencies of the R1a-B6 variants successfully expressed and characterised later in the project from library H9R3a are plotted. The lines for G1, G3 and C27 are truncated because these variants were not present in the unsorted library dataset (or in H9R1 in the case of C27). **C)** The sequence at the CDRs and surrounding residues of these R1a-B6 variants are shown aligned. Mutations are highlighted. If the mutation is one of the 5 most common mutations, it is highlighted in the corresponding colour.



C)

CDR1 and surrounding residues

	26	27	28	29	30	31	32	33	34	35
R1a-B6	G	S	F	F	S	R	Y	R	M	G
WC4	G	S	F	F	S	N	Y	R	M	G
WC11	G	S	F	F	S	I	Y	R	M	G
WC21	E	S	F	F	S	I	Y	R	M	G
WC25	G	S	F	F	S	R	Y	R	M	G

CDR2 and surrounding residues

	50	51	52	53	54	55	56	57	58	59	60	61	62	63
R1a-B6	S	I	A	Y	D	G	S	T	S	Y	A	D	P	V
WC4	S	I	T	Y	D	G	S	T	S	Y	A	D	S	A
WC11	G	I	A	Y	D	G	S	P	S	Y	A	D	P	V
WC21	S	I	A	Y	D	G	T	T	S	Y	A	D	P	V
WC25	S	I	A	Y	D	G	S	T	N	Y	A	D	P	V

CDR3

	93	94	95	96	97	98	99	101	102
R1a-B6	N	L	D	P	P	G	I	L	Y
WC4	N	L	D	P	P	G	I	L	Y
WC11	N	L	N	P	P	G	I	L	D
WC21	N	L	N	P	P	G	I	L	S
WC25	N	L	A	P	P	G	I	L	Y

Figure R2 10a,b,c: Enrichment of individual mutations and variants of interest in H9 sorted libraries H9R1, 2 and 3b A) Similar to figure R2 9a, The enrichment of individual mutations in response to H9 sorting selection is tracked with Illumina sequencing. In this figure, H9R3a is substituted for the H9R3b library which was produced by sorting the H9R2 library with an amended protocol. The frequency traces of the top 5 most frequent mutations in library H9R3b (R31I, R31S, S50R, D95N, P62S) are coloured green, orange, blue, purple, red respectively. **B)** shows the frequency of variants WC4, WC11, WC21 and WC25 throughout the selection process. **C)** The mutations in the CDR regions of these 5 variants are compared and mutations in the top 5 frequent mutations in library H9R3b are highlighted in the corresponding colour.

5. Results chapter 3: Expression and characterisation of antibody variants

5.1. Chapter introduction

The sorting experiments in the previous chapter (section 4) yielded a panel of antibodies to be further characterised. These antibodies were selected for maximum H9 antigen binding in the yeast display system. However, the objective of the project was not just to produce improvements in H9 HA binding but broaden the antigen binding range. Therefore, an optimised R1a-B6 variant needed to retain affinity comparable to R1a-B6 for H1 and H5 HA antigens as well as improvements to H9 binding activity.

When characterising and comparing a range of antibodies it is helpful to make use of more than one kind of assay. Results vary substantially between different assay types and a range of comparative assay results increases confidence that an observed improvement in antibody function is general and not specific to one assay. In addition, an antibody can be functionally characterised in different ways in different experiments. For example, SPR experiments return kinetic data on the binding interaction while PV assays indicate relative ability to neutralise virus. Finally, many influenza antigens are only available in one format (purified recombinant, whole virus antigen standard, PV, etc) so a range of experiments allows a wider range of antigens to be tested.

5.2. Expression and purification of R1a-B6 variants

All of the experiments in this chapter required purified VHH so R1a-B6 and R1a-B6 variants were expressed in *E.coli*, extracted and purified. Recombinant VHH was used in surface plasmon resonance experiments for determination of k_{on} and k_{off}, in pseudovirus neutralisation assays to assess relative potency against a virus model and in ELISAs to gather more data on binding range and relative activity.

Genes encoding R1a-B6 variants of interest were excised either from pNIBS-5 plasmids derived from monoclonal yeast colonies or from synthesised VHH genes ordered based on sequence data. VHH genes were cloned into the pNIBS-

1 plasmid and transformed into a strain of *E.coli* (either WK6 or BL21) and used for protein expression. The pNIBS-1 plasmid facilitated IPTG-induced expression of the encoded VHH in the bacterial periplasm. All expressed VHH domains included a His-tag and a Myc-tag at the C-terminal end. Making use of the His-tag, VHH domains were purified using TALON resin and the resulting eluate was buffer exchanged with PBS. Full details of the protein expression and purification protocol are in section 2.6, as well as details of quality control. Yields of purified VHH and the experiments they were used in are listed in table R3 1. Figure R3 1a,b,c,d,e,f shows Coomassie stained gels and Myc-tag western blots of purified protein samples. These gels verify the molecular weight, purity and presence of Myc epitope tags of the purified VHH samples.

VHH domain	Volume of culture (ml)	Volume of purified VHH solution (ml)	Concentration of purified VHH solution (mg/ml)	SPR (HA immobilised on chip)	PV neutralisation assays	ELISA	SPR (VHH immobilised on chip)	Flow cytometry epitope mapping experiments
R1a-B6	800	4	0.108	X	X	X	X	X
R4A	400	1	0.059	X				
G1	400	0.350	0.076	X				
G3	400	0.350	0.115	X				
C23	400	2	0.032	X				
C27	800	4	0.106	X	X			
WC4	800	1	0.264	X				
WC11	800	2	0.067	X	X			X
WC21 batch 1	800	2	0.253	X	X			X
WC25	800	2	0.200	X				
WC21 batch 2	1600	8	0.807			X	X	X
VarH	400	2.5	0.134			X	X	
VarD	400	2.5	0.910			X	X	
VarI	800	3	1.970			X		
VarJ	800	4	0.858			X	X	
VarP	800	4	0.650			X		
VarQ	800	4	1.316			X	X	

Table R3 1: Production and application of VHH for this project Listed above are the R1a-B6 variants produced and purified as part of this project and their production yields. The experiments in which the purified VHH were used are also listed.

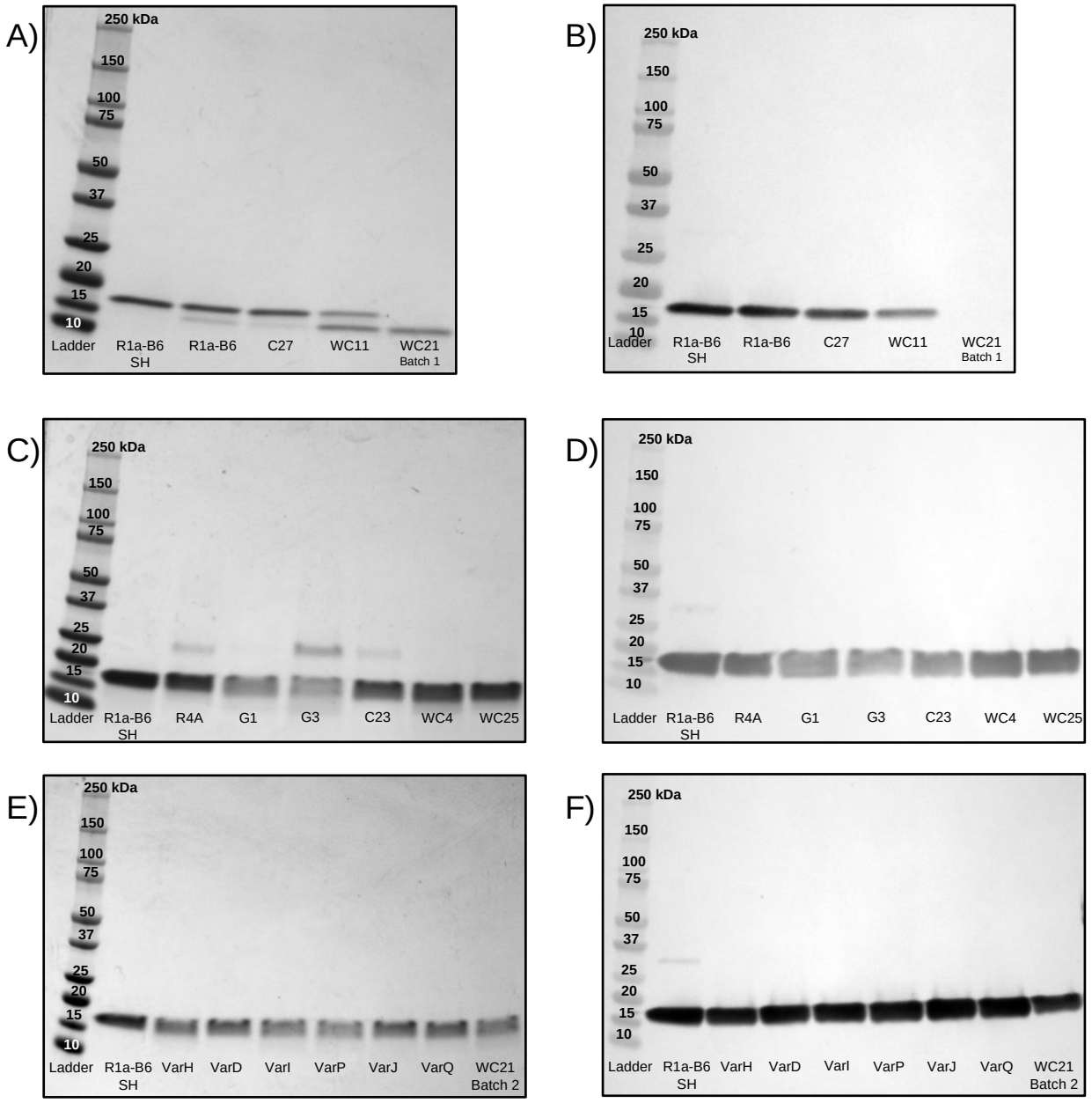


Figure R3 1a,b,c,d,e,f: Coomassie stained gels and α-Myc-tag western blots of all the VHH domains expressed in this project For each SDS PAGE gel run, purified VHH samples in PBS were diluted to the same concentration. **A)** and **B)** samples were at 0.050mg/ml, **C)** and **D)** samples were at 0.120mg/ml, **E)** and **F)** samples were at 0.134mg/ml. Gels **A)**, **C)** and **E)** were stained with Coomassie dye while **B)**, **D)** and **E)** show western blots made from the SDS gels using an α-Myc-tag antibody. All gels contained a sample of R1a-B6 with a Myc epitope tag purified prior to the project as a positive control marked "R1a-B6 SH" on the images. The Coomassie stained gels and western blots confirm that the samples, for the most part, contained approximately equal quantity of a Myc-tagged protein identical in size to the "R1a-B6 SH" (~17kDa) with minimal extra bands. R4A and C23 contain faint unidentified extra bands and G3 contains substantial unidentified extra bands. A recurrent issue of the expression pipeline used is the loss of the Myc-tag from the expressed protein. This does not appear to have a functional effect on the ability of the VHH domain to bind to antigen. Faint bands slightly below the main bands of R1a-B6 and C27 appear to show VHH without the tag. WC11 has a more substantial Myc-tag negative band and WC21 batch 1 appears to have lost the tag completely.

Of the clones selected for expression in results chapter 2, some failed to express in soluble form at all or expressed in insufficient quantity to collect useful results. These clones were G2, G4 and C14. In addition, WC11, one of the two most promising clones, consistently expressed poorly and only 1 batch was created without significant extra unidentified bands of protein. As a result, WC11 was omitted from some experiments when supplies of the successfully purified batch ran low.

5.3. Screening of antibodies using surface plasmon resonance

With the relatively large number of antibodies to be characterised, a rapid and non-resource intensive method was required to assess the panel. Surface plasmon resonance using HA-coated chips was selected for this initial characterisation. Surface plasmon resonance involves immobilisation of one interacting molecule, the ligand (in this case HA), to a gold film. Light is reflected off the film at an angle dependent on the effective mass of the film. Effective mass includes anything attached to the film and to the immobilised ligand. Very small changes in effective mass can be measured in real time based on the angle of the reflected light. By passing a solution containing the other interacting molecule, the analyte (in this case VHH), the binding interactions between the immobilised molecule and the molecule in solution can be detected by changes in the effective mass of the membrane[149, 211]. In all SPR experiments in this project EDC/NHS ester immobilisation of the ligand was used[212].

An SPR protocol called single cycle kinetics (SCK) was used in this set of experiments[150]. In the first phase (binding phase) of SCK, a range of solutions of the same analyte are passed over the ligand in order of increasing concentration. In the subsequent dissociation phase, buffer without analyte is passed over the ligand, allowing the analyte to exponentially dissociate. Throughout the process, the change in effective mass of the SPR film is monitored and the change in mass of bound analyte over time allows estimation of k_{on} and k_{off}.

Many SPR protocols, including SCK protocols, include a regeneration step. Regeneration steps induce the rapid dissociation of analyte from ligand by a

change in the solution passing over the ligand. However, they can also cause degradation of the ligand[212]. Influenza HA has previously been observed to degrade rapidly from regeneration cycles and was in limited supply so regeneration cycles were omitted in this experiment. Instead, the dissociation phase was extended to allow more analyte to dissociate from the SPR surface. The disadvantage of this method is that some analyte remains bound to the ligand surface while the next analyte is being tested, reducing the number of available binding sites on the ligand.

Figure R3 2 compares the SPR traces recorded on a chip coated with H9 HA (A/Hong Kong/1073/1999) of R1a-B6 and WC11 (a variant which was measured to have an increased k_{on} and a decreased k_{off} compared to the unaltered antibody). At the higher concentrations of VHH, the amount of WC11 starts to plateau, indicating that almost all the available HA epitopes on the chip are occupied. In contrast, in the R1a-B6 run, the amount of VHH binding to the chip is still rising just before the dissociation phase starts. This difference is primarily due to a difference in k_{on}. k_{off} differences mostly manifest in the dissociation phase. Both dissociation traces show an exponential dissociation of VHH but the gradient of the R1a-B6 curve is far steeper at any given point because the VHH is dissociating at a much greater rate. At the end of the 4000 second dissociation phase (figure does not show full dissociation phase), the level of R1a-B6 bound to the chip had dropped by 89% compared to its maximum while the WC11 levels had only dropped by 31%. This difference is entirely due to difference in off rate[211, 213].

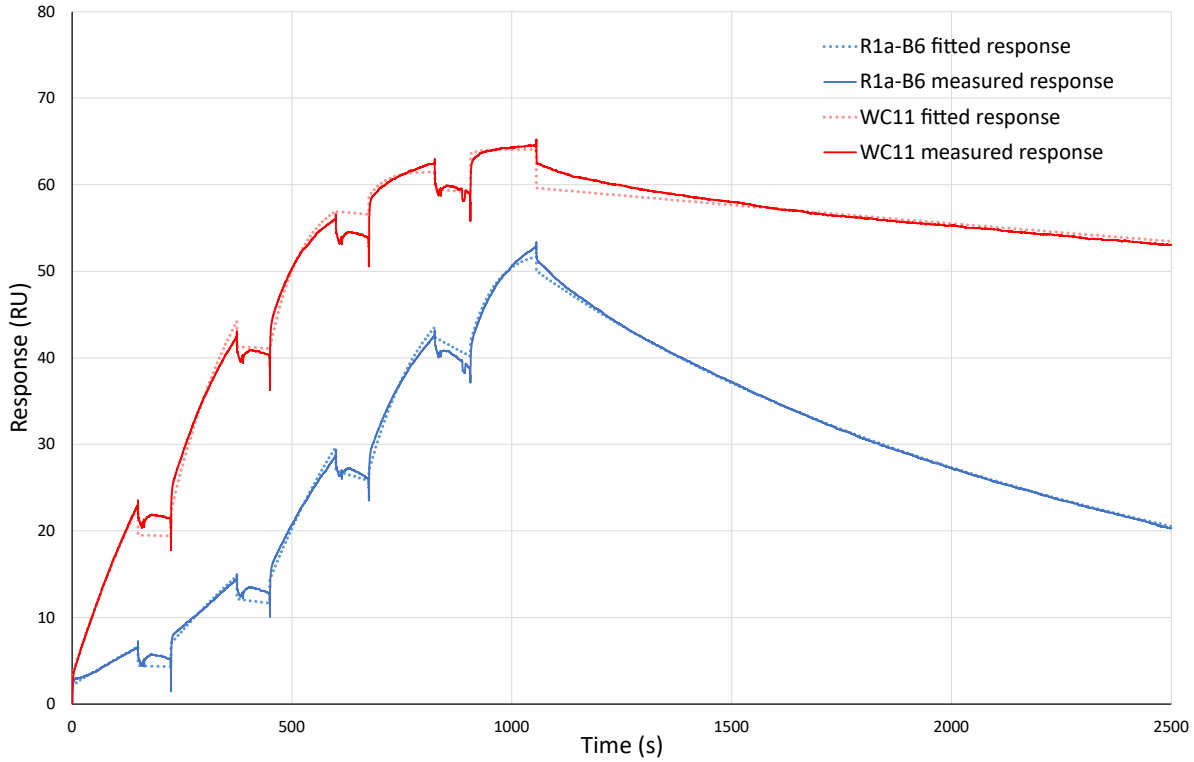
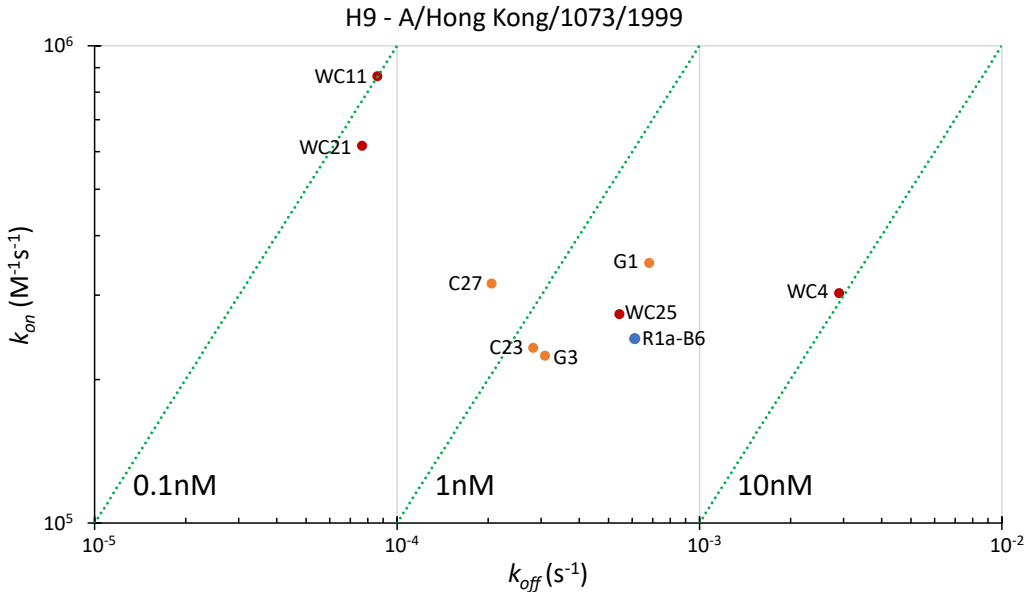


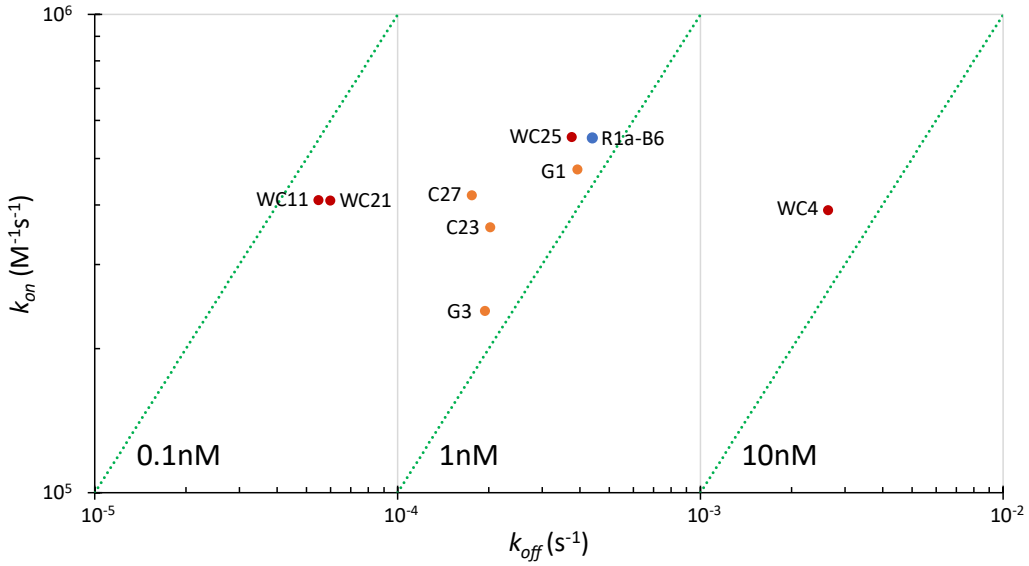
Figure R3 2: Comparison of R1a-B6 and WC11 interacting with an H9 coated SPR chip Response (measured in arbitrary response units) represents the change in effective mass of the ligand-coated surface. WC11 is measured to have a higher k_{on} and a slower k_{off} compared to unmodified R1a-B6. The differences in both parameters contribute to the faster binding of WC11 during the binding phase but only the k_{off} of the two antibodies determines the rate at which they dissociate from the chip.

Surface plasmon resonance chips were coated in H9, H5 and H1 (A/Hong Kong/1073/1999, A/Anhui/1/2005 and A/California/7/2009) recombinant HA antigen and SCK was used to analyse the binding interactions of the R1a-B6 variants with the antigens. The measured values of k_{on} and k_{off} are plotted in figure R3 3a,b,c and are tabulated with SEM in table R3 2a,b,c. The R1a-B6 variant R4A, which showed no detectable binding to any of the antigens, was omitted from the figures and tables.

A)


H9 - A/Hong Kong/1073/1999
WC11
WC21
C27
G1
WC4
WC25
R1a-B6
C23
G3
0.1nM
1nM
10nM
10^6
10^5
10^{-5}
10^{-4}
10^{-3}
10^{-2}
k_{on} ($M^{-1}s^{-1}$)
k_{off} (s^{-1})

B)


H5 - A/Anhui/1/2005
WC25
R1a-B6
G1
WC11
WC21
C27
C23
WC4
G3
0.1nM
1nM
10nM
10^6
10^5
10^{-5}
10^{-4}
10^{-3}
10^{-2}
k_{on} ($M^{-1}s^{-1}$)
k_{off} (s^{-1})

C)

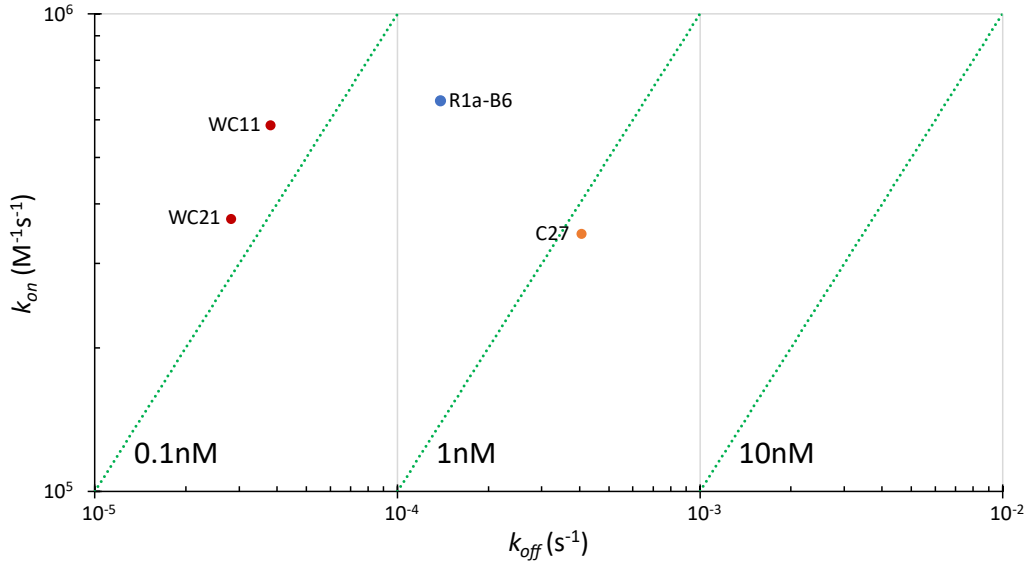

H1 - A/California/7/2009
R1a-B6
WC11
WC21
C27
0.1nM
1nM
10nM
10^6
10^5
10^{-5}
10^{-4}
10^{-3}
10^{-2}
k_{on} ($M^{-1}s^{-1}$)
k_{off} (s^{-1})

Figure R3 3a,b,c: Kinetic parameters of R1a-B6 and R1a-B6 variants shown in affinity plots k_{on} and k_{off} values, as measured by surface plasmon resonance using **A)** H9, **B)** H5 and **C)** H1 recombinant HA coated chips, are plotted against each other for a panel of purified VHH. In each plot, points are labelled with the antibody name and coloured according to the variant library of origin (orange for H9R3a and red for H9R3b). The point corresponding to R1a-B6 is coloured blue. The points plotted are a mean of 2-6 measurements of each kinetic parameter (see table R3 2a,b,c). Dotted green lines on the graph are isoaffinity diagonals. Due to the relationship between k_{on}, k_{off} and KD, each point along those lines represents the same affinity (annotated at the bottom of the plot area).

A)

H9 - A/Hong Kong/1073/1999	k_{on} x10^5($M^{-1}s^{-1}$) ±SEM	k_{off} x10^{-5}(s^{-1}) ±SEM	KD (nM)	n
R1a-B6	2.43 ± 0.1	61.0 ± 0.9	2.51	6
C27	3.19 ± 0.06	20.5 ± 0.6	0.644	6
G1	3.52 ± 0.2	68.1 ± 1	1.94	6
G3	2.25 ± 0.3	30.8 ± 3	1.37	6
C23	2.33 ± 0.03	28.2 ± 0.2	1.21	6
WC4	3.03 ± 0.07	290 ± 10	9.54	6
WC11	8.64 ± 0.1	8.60 ± 0.4	0.0995	6
WC21	6.19 ± 0.1	7.66 ± 0.2	0.124	6
WC25	2.74 ± 0.1	54.2 ± 0.9	1.98	2

B)

H5 - A/Anhui/1/2005	k_{on} x10^5($M^{-1}s^{-1}$) ±SEM	k_{off} x10^{-5}(s^{-1}) ±SEM	KD (nM)	n
R1a-B6	5.52 ± 0.2	44.0 ± 3	0.798	6
C27	4.19 ± 0.1	17.6 ± 1	0.420	6
G1	4.75 ± 0.2	39.2 ± 1	0.809	6
G3	2.40 ± 0.08	19.4 ± 1	0.825	6
C23	3.59 ± 0.1	20.2 ± 0.5	0.563	6
WC4	3.90 ± 0.1	264 ± 6	6.76	4
WC11	4.09 ± 0.1	5.48 ± 0.2	0.134	6
WC21	4.08 ± 0.5	5.98 ± 0.6	0.146	5
WC25	5.55 ± 0.3	37.6 ± 2	0.677	4

C)

H1 - A/California/07/2009	k_{on} x10^5($M^{-1}s^{-1}$) ±SEM	k_{off} x10^{-5}(s^{-1}) ±SEM	KD (nM)	n
R1a-B6	6.57 ± 0.08	13.4 ± 0.06	0.211	4
C27	3.46 ± 0.2	40.5 ± 0.4	1.17	4
WC11	5.84 ± 0.2	3.80 ± 0.07	0.0650	2
WC21	3.72 ± 0.1	2.81 ± 0	0.0755	2

Table R3 2a,b,c: Kinetic parameters of R1a-B6 and variants measured by surface plasmon resonance on HA coated chips k_{on} and k_{off} values (3 significant figures) listed are mean of 2-6 measurements ± the standard error of the mean (1 significant figure). The number of measurements contributing to each average value is listed in the rightmost column labelled n. KD was calculated by dividing mean k_{off} by mean k_{on} and is listed to 3 significant figures

Unsurprisingly, most of the selected antibodies show increased affinity for H9 antigen, although for most antibodies tested, the affinity improvements were modest, 1.2 to 2-fold increases. WC4 was the outlier of this set with a 4-fold affinity decrease. Of the antibodies selected from the H9 round 3a library, C27 gained the most in affinity with an approximately 4-fold affinity increase. More significant though were the affinity increases displayed by WC11 and WC21 from the improved H9 round 3b library which showed 25 and 20-fold improvements to affinity respectively.

The H5 affinities showed a similar picture. Like with H9, WC25, G1, G3 and C23 showed similar affinity to R1a-B6 or slight improvements. C27 had the highest affinity of the H9 round 3a library variants at 2-fold improvement over R1a-B6 and higher still were WC11 and WC21 with 5 to 6-fold KD improvement. However, while several of the variants showed on rate improvements with H9, none did with H5 and most showed reduced on rate. Improvement to H5 affinity came only from off rate.

The H1 antigen was tested with a reduced set of VHH variants. As with H5, there was no k_{on} improvement observed, all antibodies tested showed lower on rates than R1a-B6. Despite their lower on rates, WC11 and WC21 showed improved affinity compared to the unaltered VHH due to much lower off rates. Unlike with H9 and H5, C27, had a lower affinity than R1a-B6 for H1 antigen.

Of the antibodies characterised, the clear stand out of improved antibodies are WC11 and WC21 which showed affinity improvements against all 3 antigens tested with particularly dramatic improvement to H9 affinity. The trade off to these improvements seems to be modest loss of on rate against H5 and H1 antigen. C27 also showed binding breadth improvements but more modest than those of WC11 and WC21.

The protocol used in these experiments was limited by the lack of a suitable regeneration cycle between runs. This meant that chips remained contaminated with antibodies previously tested. In some cases, subsequent results were distorted, especially after the HA ligand on chips was exposed to high affinity VHH. In the case of R1a-B6 and WC11 on the H9 chip for example, by the end of the runs depicted in figure R3 2, the level of R1a-B6 on the chip dropped to 11% of its maximum value, leaving a large majority of stem epitope sites available for binding in the next SCK run. Thus, the distortion of subsequent results was minimal. However, the same chip retained 69% of the maximum level of bound WC11 at the end on the run, severely limiting the number of available epitope sites for the next run. In cases such as these, subsequent runs showed decreased on rates and increased off rates and often these contaminated runs had to be discarded from analysis.

Despite the accuracy limitations imposed by the influence of residual antibody on the SPR chips, this set of SPR data served its function to identify the most promising antibodies to characterise in more resource and time intensive experiments. Later in this chapter, SPR experiments using a different protocol without the same limitation were used to characterise the binding of the lead antibody WC21 to recombinant HA in greater detail.

5.4. PV neutralisation data

The next step in antibody characterisation was to test how the differences in antibody affinity translated to neutralisation potency. Of the panel characterised in the previous section, only a few could be picked for pseudovirus neuralisation assays due to time constraints and amount of available pseudovirus. WC11, WC21 and C27 were picked to be compared in the assay to R1a-B6 as these variants had the highest H9 affinity as well as retaining or improving on H5 and H1 affinity. R1a-G2 was selected as a negative control[102].

Pseudoviruses are artificial constructs which model the cytosol entry mechanisms of live viruses without the ability to complete a replication cycle and, therefore, can be handled more safely and with fewer precautions compared to live virus[214, 215]. PV neutralisation titres have historically shown strong correlation with live virus titres[216]. The PV used in these experiments (kindly supplied by Joanne Del Rosario and Nigel Temperton), unlike influenza viruses, only carry HA on their envelopes. The core of these PV is similar to a lentiviral core, carrying two identical RNA molecules packaged in lentiviral capsid protein. The RNAs in these PV encode a reporter gene, firefly luciferase, for quantification of PV entry into the cell cytosol. PV were used to infect cultured human embryonic kidney cells (HEK293 cells) which, upon PV entry, would express the reporter gene. Neutralised PV are inhibited from cell entry and therefore neutralising antibodies inhibit reporter gene expression[151, 152]. PV structure and entry into a cell are summarised in figure R3 4a,b.

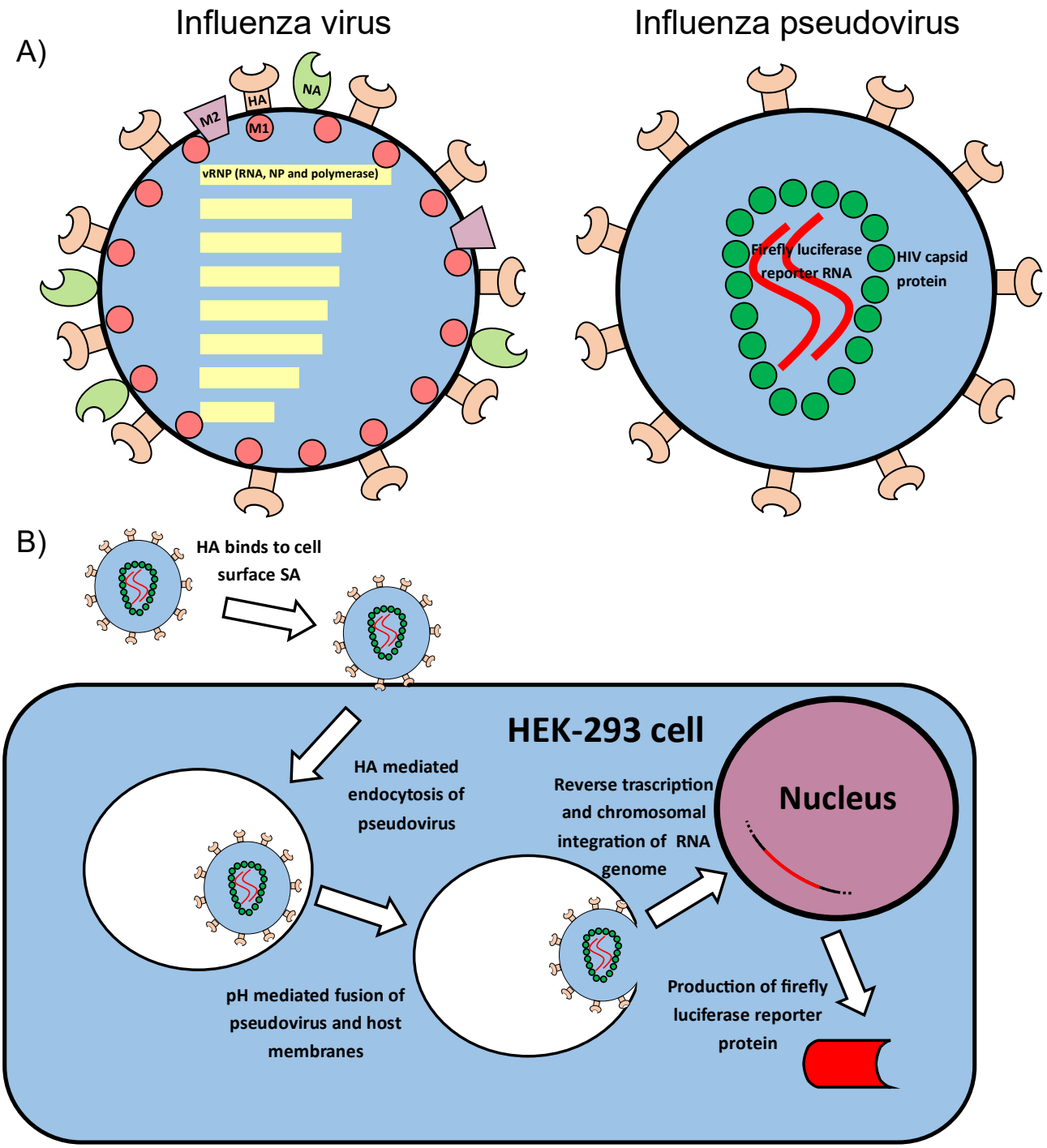


Figure R3 4a,b: Comparison of influenza virus with influenza pseudovirus A) Cartoon representations of an influenza virus and an influenza pseudovirus with a lentiviral core. While the live virus carries 3 proteins on its envelope, HA, NA and M2, only HA is necessary for viral entry into host cells so the pseudovirus only carries HA. Likewise, the 8 vRNPs of influenza and M1 are required for the complete influenza life cycle but not virion entry so are absent in the pseudovirus. Instead, the pseudovirus envelope contains an RNA encoding a reporter gene (firefly luciferase) surrounded by HIV capsid protein[215]. **B)** Cartoon representation of influenza pseudovirus entry into a host cell. The influenza pseudovirus serves as a non-pathogenic model for the entry steps of the virus life cycle. Like live influenza virus, pseudoviruses are endocytosed. Fusion with the host membrane is triggered by endosome acidification, all of which is mediated by HA. After membrane fusion, however, the HIV-like core of the pseudovirus promotes the expression of firefly luciferase by reverse transcription and chromosomal integration of the RNA genome. This allows convenient quantification of pseudovirus infection of cells in culture[151, 152].

R1a-B6, C27, WC11, WC21 and R1a-G2 were tested in triplicate against PV expressing HA from H1, H2, H5, H7 and H9 viruses (H1N1 – A/England/195/2009, H2N2 – A/Korea/426/1968, H2N2 – A/England/1/1966, H5N1 – A/Vietnam/1203/2004, H7N9 – A/Shanghai/2/2013, H9N2 – A/Hong Kong/1073/1999, H9N2 – A/Chicken/Israel/291417/ 2017).

Serial dilutions of antibodies were incubated for an hour with approximately 10^6 relative luminescence units (RLU – a measure of the luminescence given out by cells infected with PV and treated with a luciferase substrate) of PV before 1.5 x 10^4 HEK cells were added to the incubated mixtures. Cells were incubated at 37°C for 48 hours before the supernatant was drained and the luminescence of the cells was measured. The serial dilution spanned from 256nM to 0.5pM in 20 2-fold steps (for the H7 PV, the range was 1024nM-0.5pM in 22 2-fold steps).

Measured luminescence values were inverted and normalised to give percent pseudovirus inhibition. 100% inhibition was defined as the mean luminescence of wells containing HEK293 cells with no PV. 0% inhibition was defined as the mean luminescence of wells with cells and PV but no antibody. By fitting sigmoidal curves to the % inhibition values across a dilution series, the concentration at which the entry of PV was 50% inhibited (IC50) was estimated[152, 217]. Fitted PV neutralisation curves are shown in figure R3 5a,b,c,d,e,f,g,h and tables showing IC50 values and IC50 fold potency change compared to R1a-B6 are shown in table R3 3a,b.

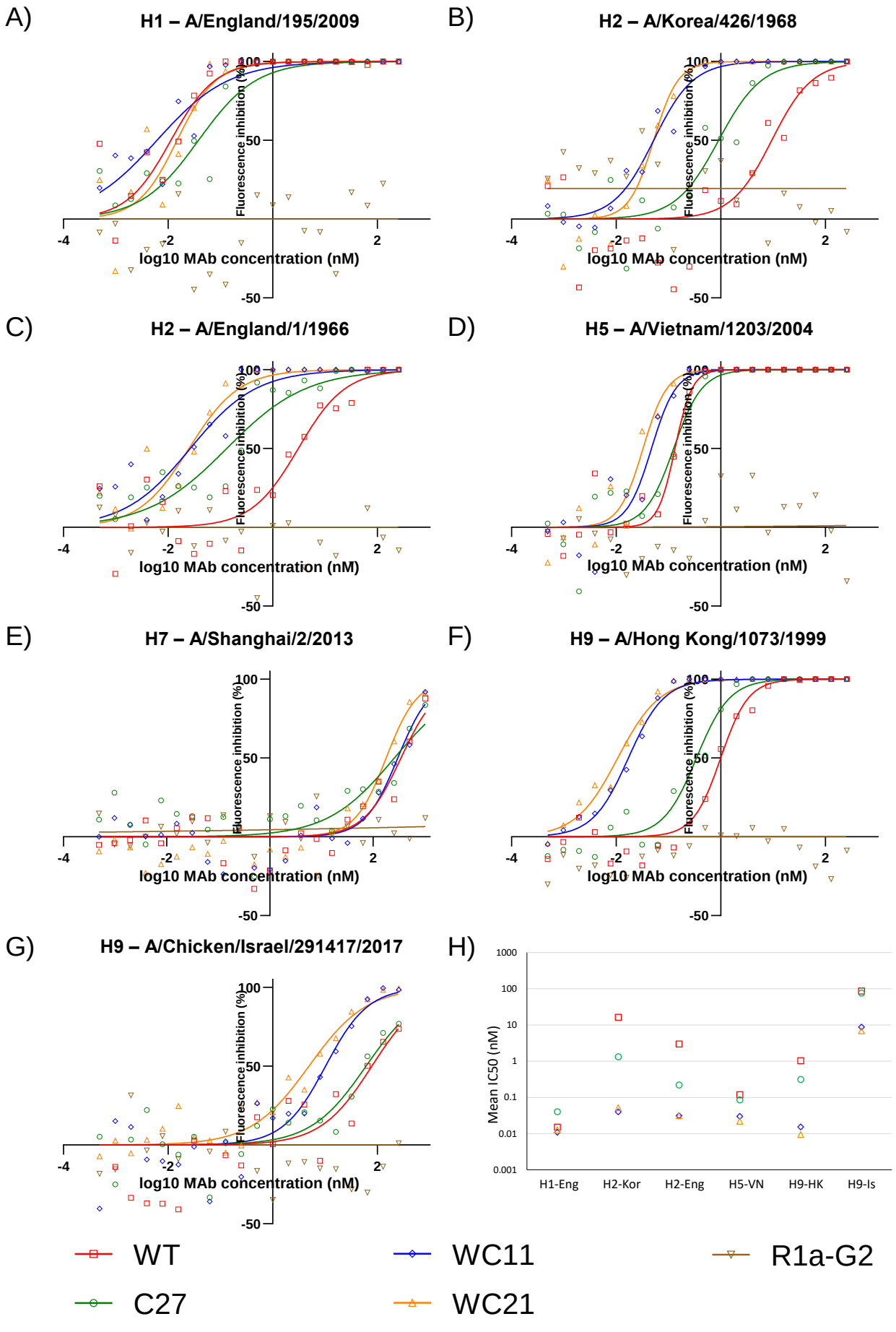

A)
H1 – A/England/195/2009
B)
H2 – A/Korea/426/1968
C)
H2 – A/England/1/1966
D)
H5 – A/Vietnam/1203/2004
E)
H7 – A/Shanghai/2/2013
F)
H9 – A/Hong Kong/1073/1999
G)
H9 – A/Chicken/Israel/291417/2017
H)
Fluorescence inhibition (%)
log10 MAb concentration (nM)
Mean IC50 (nM)
H1-Eng
H2-Kor
H2-Eng
H5-VN
H9-HK
H9-Is
WT
WC11
R1a-G2
C27
WC21

Figure R3 5a,b,c,d,e,f,g,h: Pseudovirus neutralisation curves to determine IC50s of R1a-B6 and variants Lentivirus-based pseudoviruses (PV) were used as a model to approximate live virus neutralisation. R1a-B6 variants C27, WC11 and WC21 were compared to the unaltered antibody in their ability to neutralise PV of the subtypes **A)** H1, **B)** and **C)** H2, **D)** H5, **E)** H7, **F)** and **G)** H9. R1a-G2, a VHH antibody with no anti-HA activity was used as a negative control. Fluorescence measurements of HEK-293 cells were taken 48 hours after they had been exposed to PV premixed with antibody. High fluorescence signal indicated that the cells had been successfully infected by the PV and low signal indicated that the infection had been inhibited. Fluorescence data was normalised by defining the signal from wells with PV and no antibody as 0% inhibited and cells with no antibody and no PV as 100% inhibited (meaning that high fluorescence measurements would be normalised to low percentage inhibition values and vice versa). Each point on the graphs represents the mean average of 2 wells and a sigmoid curve was fitted to each dilution series. The further left a curve is, the lower the concentration of antibody required to neutralise the PV. IC50, the concentration at which the fitted neutralisation curve crossed 50% fluorescence inhibition was used as a measure of neutralisation potency[152]. **H)** For each antibody-PV combination, the mean of 3 estimates of IC50 is plotted against the PV used to attain the measurements to compare the potency profile of each R1a-B6 variant.

A

PV IC50 ± SEM (nM)	R1a-B6	WC11	WC21	C27	R1a-G2 (negative control)
H1N1 - A/England/195/2009	0.0152 ± 0.007	0.0109 ± 0.006	0.0127 ± 0.005	0.0406 ± 0.002	>256
H2N2 - A/Korea/426/1968	16.4 ± 5	0.0402 ± 0.01	0.0516 ± 0.01	1.33 ± 0.2	>256
H2N2 - A/England/1/1966	3.01 ± 0.7	0.0311 ± 0.003	0.0313 ± 0.006	0.223 ± 0.07	>256
H5N1 - A/Vietnam/1203/2004	0.120 ± 0.01	0.0301 ± 0.009	0.0220 ± 0.006	0.0859 ± 0.02	>256
H7N9 - A/Shanghai/2/2013	>256	>256	>256	>256	>256
H9N2 - A/Hong Kong/1073/1999	1.04 ± 0.2	0.0153 ± 0.003	0.00947 ± 0.001	0.314 ± 0.07	>256
H9N2 - A/Chicken/Israel/2191417/2017	87.4 ± 20	8.81 ± 2	6.92 ± 1	76.5 ± 30	>256

B

Fold change compared to R1a-B6	WC11	WC21	C27
H1N1 - A/England/195/2009	1.4	1.2	0.37
H2N2 - A/Korea/426/1968	410	320	12
H2N2 - A/England/1/1966	97	95	14
H5N1 - A/Vietnam/1203/2004	4.0	5.5	1.4
H9N2 - A/Hong Kong/1073/1999	68	110	3.3
H9N2 - A/Chicken/Israel/2191417/2017	9.9	13	1.1

Table R3 3a,b: Pseudovirus neutralisation IC50 data Sigmoid curves fitted to normalised fluorescence inhibition values (see figure R3 5a,b,c,d,e,f,g,h) were used to estimate the concentration of antibody required to reduce the pseudovirus fluorescence signal by 50%. Each antibody was tested against each PV in triplicate. The values recorded in **A)** are means accompanied by the standard error of mean (SEM) of the 3 estimated IC50 values, both in nM.

Mean values are recorded to 3 significant figures and SEM is recorded to 1 significant figure. **B)** The fold change in mean IC50 of R1a-B6 variants compared to unaltered R1a-B6 to 2 significant figures. H7 PV results are omitted because IC50 values were highly variable. Fold improvements greater than 20 are highlighted in green.

All the R1a-B6 variants tested neutralised PV expressing H9 - A/Hong Kong/1073/1999 more potently than the unaltered R1a-B6. However, in line with SPR kinetics measurements on the same antigen, the potency improvement for WC11 and WC21 (68 and 110-fold respectively) was much larger than the potency improvement seen for C27 (3.3-fold).

For the other H9 PV tested (A/Chicken/Israel/2191417/2017), potency for every antibody was much lower compared to potency against H9 - A/Hong Kong/1073/1999. Nonetheless, WC11 and WC21 displayed enhanced potency compared to R1a-B6 and C27.

Following the trend of SPR affinity measurements, WC11 and WC21 showed modest IC50 improvements against both H1 and H5 PV compared to R1a-B6. C27 PV IC50 values also mirrored SPR affinity measurements. C27 was less potent in neutralising H1 PV compared to R1a-B6 and slightly more potent against H5 PV.

Surprisingly, the most substantial increases in potency detected were in H2 neutralising titre. Both H2 PV showed similar trends with 95 to 410-fold increases in WC11 and WC21 potency compared to R1a-B6 and 12 to 14-fold potency increases for C27 compared to R1a-B6. These potency increases were somewhat unexpected as none of these antibodies had been selected by yeast display sorting for binding to H2 antigen and the H2 binding selection experiments seemed to yield no antibodies which would bind to H2 at a level detectable on the flow cytometer.

The potency of parental antibody, R1a-B6, varies a great deal depending on the influenza A group 1 PV subtype (R1a-B6 is 1100-fold more potent against H1 – England than against H2 - Korea). In comparison, the inter subtype antigenic differences between group 1 HAs seem to have little effect on the potency of WC11 and WC21 (except for H9 – Israel, an influenza virus isolate from a chicken). The potencies of WC11 for H1 – England, H2 – England, H2 – Korea

and H9 – Hong Kong are all within 4-fold of each other. Similarly, for WC21, potencies for the same set of PV are within 5-fold. Clearly the substitutions distinguishing WC11 and WC21 from R1a-B6 have almost entirely overcome the barriers to high potency neutralisation of H2 and H9 PV.

Figure R3 5e shows H7 neutralisation curves. It appears that R1a-B6 and all of its variants have some neutralising activity against this PV at very high antibody concentrations in agreement with Joanne Del Rosario's observations with R1a-B6[unpublished data]. However, the calculated IC50 values were highly variable and average IC50s were all greater than 256nM so they were not listed on table R3 3a. Furthermore, there were no consistent potency differences between R1a-B6 and any of the variants on H7 PV tested so the phenomenon was not investigated further.

5.5. Identification of substitutions responsible for variant potency improvements

The data from this chapter so far has demonstrated dramatic affinity and potency improvements by the R1a-B6 variants WC11 and WC21 compared to the unaltered antibody. WC11 and WC21 are modestly more potent against H1 and H5 PV and show much larger improvements against H2 and H9 PV. Interestingly, variants WC11 and WC21 have very similar performance to each other. Across all influenza antigen measurements in SPR and PV, there is never a difference between WC11 and WC21 which is more than 2-fold. This remarkable correlation could be suggestive of a common mechanism driving the affinity and potency improvements relative to unaltered R1a-B6.

The sequences of WC11 and WC21 were compared to R1a-B6 and C27 (which also demonstrated H9 and H2 PV potency improvement) in figure R3 6a,b. WC11 and WC21 share 2 substitutions in common: R31I and D95N. In addition, both VHH domains have substitutions at position 102, WC11 had Y102D and WC21 had Y102S. C27 had substitutions at 2 out of these 3 positions, R31T and D95N. If there is a common mechanism behind the improvements seen in WC11 and WC21, it would follow this may be caused by substitutions or substitution positions in common. It was hypothesised therefore, that the broadening of the

binding and neutralisation ability of WC11 and WC21 was primarily caused by 3 key substitutions R31I, D95N and Y102D/S. Other substitutions in the sequences of WC11 (S50G and T57P) and WC21 (G26E and S56T) were hypothesised to be passenger mutations contributing little or nothing to reactivity breadth.

To investigate this hypothesis, previous yeast display binding data was re-examined. All 3 positions of interest were investigated in the alanine scan experiment (section 3.2) and the substitution D95N was investigated as it was found in the R1a-B6 clonal lineage (section 3.3) (figure R3 6c). The R31A substitution appears to enhance H5 and H9 binding at the cost of diminished H1 binding. D95A and D95N substitutions both enhanced binding to all 3 subtypes while the Y102A substitution appears to be neutral or negative. The yeast data seems to support the hypothesis that substitutions at positions 31 and 95 can enhance H5 and H9 binding. Support is stronger for position 95 as the exact substitution found in WC11 and WC21 was tested rather than a different substitution at the same position.

Library sorting Illumina sequencing data also suggested that some of these substitutions may be advantageous to H9 binding. Substitutions R31I, D95N and Y102S were all substantially enriched between the unsorted library and sorted library H9R3b (enrichment factors 4.9, 4.4 and 2.9 respectively). Contrary to this hypothesis however, substitution Y102D was depleted by 36% (enrichment factor 0.64). The hypothesised passenger mutations had frequency changes ranging from substantial depletion to modest enrichment (S50G – EF 0.1, T57P – 0.3, G26E – 1.4, S56T – 1.9).

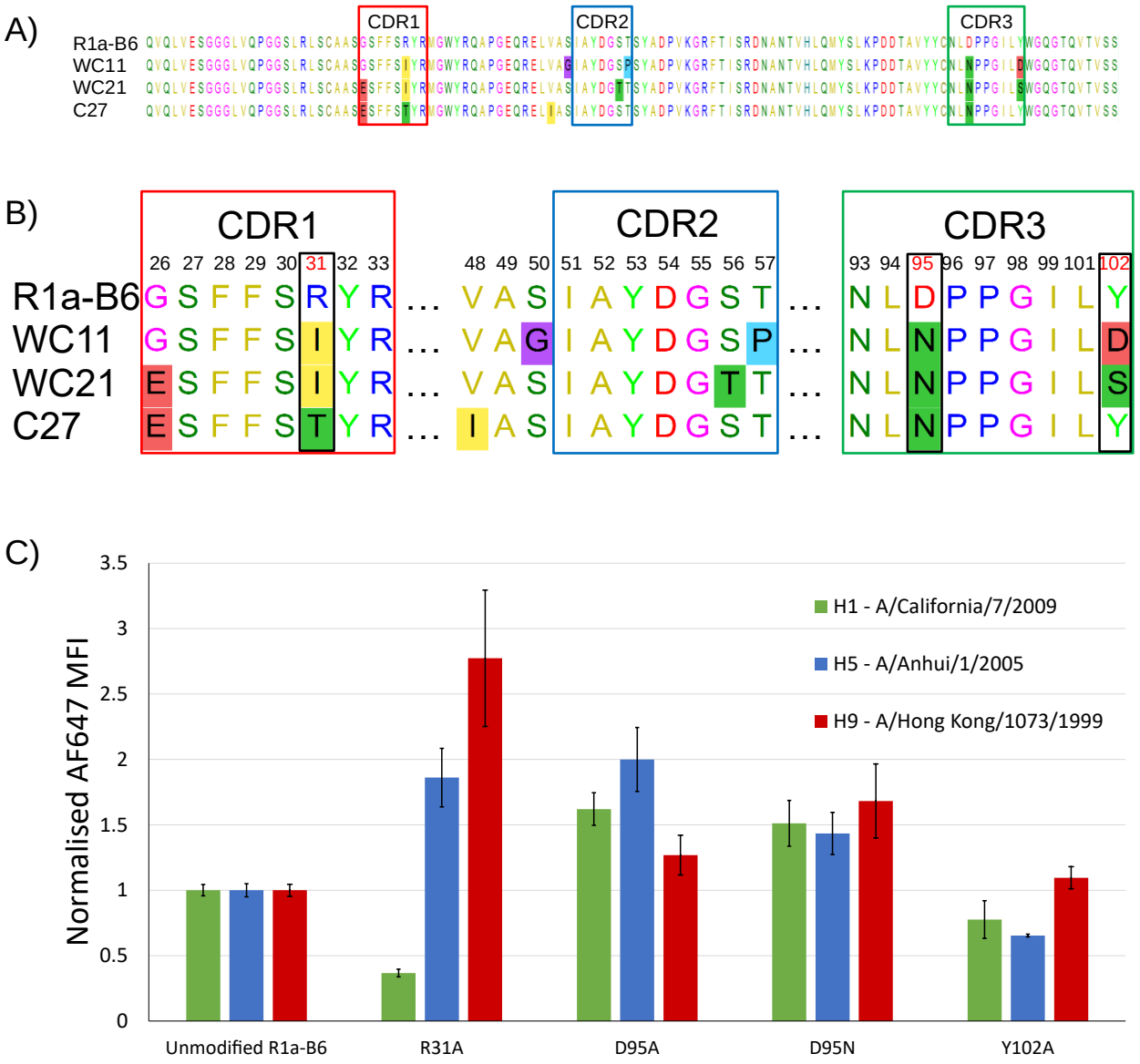


Figure R3 6a,b,c: Substitutions in R1a-B6 variants with enhanced H2 and H9 PV neutralisation A) Alignment of R1a-B6 with variants WC11, WC21 and C27 all of which have shown enhanced neutralisation potency against H9 and H2 PV. Residues differing from R1a-B6 are highlighted and the CDR regions are labelled. **B)** Alignment of the CDR regions of the same VHH domains (plus a small portion of FR2). Residues are labelled with Kabat numbering. Of the antibodies with improved H9 neutralisation, all 3 have substitutions at positions 31 and 95 and the two antibodies with the greatest potency improvements, WC11 and WC21, both have substitutions at position 102 as well. **C)** The effects of previously characterised substitutions at positions 31, 95 and 102 on yeast display HA binding. Mutant antibodies are compared to unmodified R1a-B6. As previously, mean MFI values are normalised to mean MFI for unmodified R1a-B6 and error bars show SEM. Results in this figure first displayed in figure R1 2i and figure R1 5c.

To investigate the 3 key substitutions hypothesis further, a new set of VHH domains were expressed and purified containing substitutions at positions 31, 95 and 102. The VHH domains had the same sequence as R1a-B6 with the following

substitutions: VarH (R31I), VarD (D95N), VarI (Y102D), VarJ (Y102S), VarP (R31I, D95N, Y102D), VarQ (R31I, D95N, Y102S) (see table R3 1 and figure R3 1e,f). This set of new VHH domains allowed the effect of each substitution at positions 31,95 and 102 to be individually examined as well as the effect of the trio of substitutions found in WC11 without the two hypothesised passenger mutations (S50G and T57P) and the trio of substitutions found in WC21 without the two hypothesised passenger mutations (G26E and S56T).

5.6. Enzyme-linked immunosorbent assays (ELISAs)

In order to gather further data comparing R1a-B6 binding to WC21 binding as well as the 3 key substitution hypothesis, ELISAs were carried out examining binding of VHH to whole virus influenza antigen.

ELISAs have many possible setups but the ELISA protocol used in this study was an indirect ELISA. Briefly, freeze-dried influenza antigen standards of subtypes H1, H2, H3, H5, H7, H9 (NIBSC, A/California/7/2009, A/Singapore/1/1957, A/Texas/50/2012, A/Vietnam/1194/2004, A/Anhui/1/2013, A/Hong Kong/1073/1999) were reconstituted in PBS and used to coat adsorptive 96-well plates.

Plates were incubated with dilution series of VHH antibodies from 30 to 0.0003µg/ml in 11 √10-fold steps (a 10-fold dilution every 2 steps). The antibodies used were R1a-B6, WC21, VarH, VarD, VarI, VarJ, VarP, VarQ and R1a-G2 (negative control). In addition, NB3-08 and NB7-09[123] were used as positive controls for H3 and H7 antigen respectively. Unfortunately, production of a new batch of WC11 failed owing to low expression of the antibody and not enough VHH was available from the original batch for this experiment.

Plates were subsequently incubated with an α-Myc-tag antibody fused to horseradish peroxidase (HRP). The antibody fusion bound to the remaining VHH which was in turn bound to antigen. ELISA plates were developed with a HRP substrate and the reaction was quenched with acid. HRP-catalysed substrate cleavage resulted in increased 450nm wavelength light absorption. Therefore, the 450nm absorption value (A450) in each well in the 96-well plate corresponded

to the amount of remaining bound antibody shortly before addition of HRP substrate. Figure R3 7 summarises the mechanism of the type of ELISA used in this project.

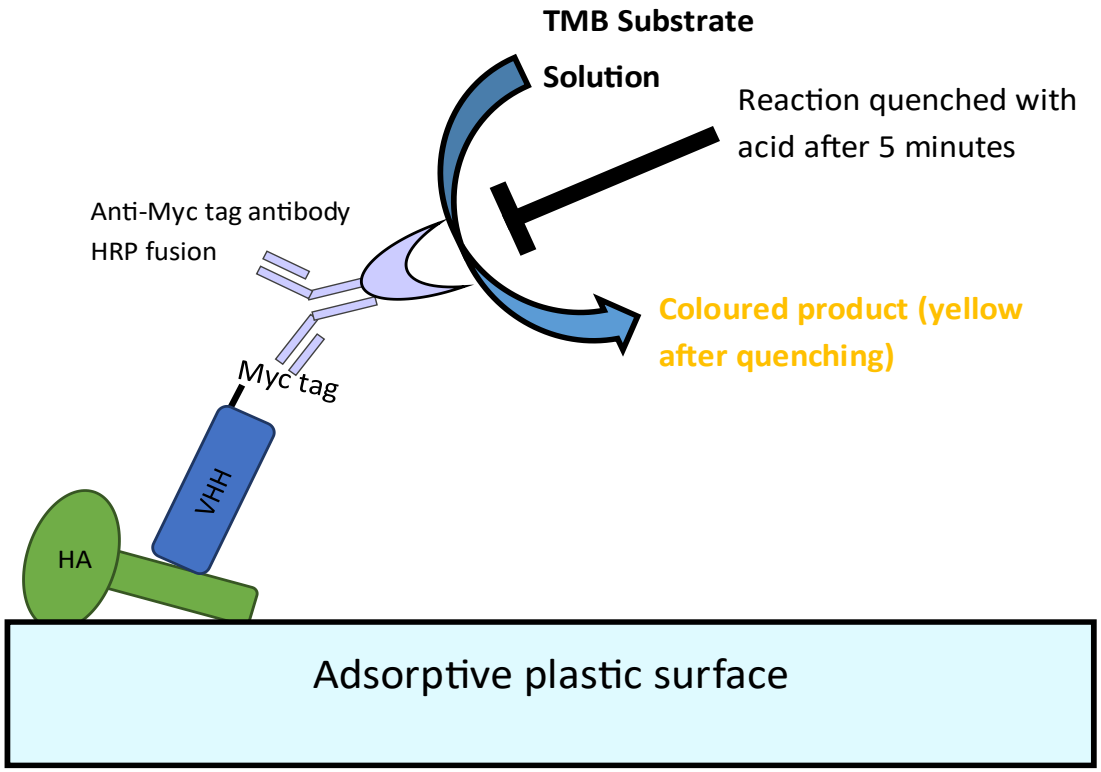


Figure R3 7 Cartoon diagram of the mechanism of the ELISA experiments used in this project Overnight incubation of adsorptive 96 well plates with reconstituted antigen standards caused adsorption of viral protein to the plate surface. Solutions of Myc-tagged VHH and anti-Myc-tag antibody HRP fusion were serially incubated in the wells to bind VHH to HA and HRP fusion to VHH. The addition of a 3,3′,5,5′-Tetramethylbenzidine (TMB) substrate solution allowed bound HRP to be quantified by the production of coloured products catalysed by the enzyme. Quenching with sulfuric acid halted the reaction and turned reaction products yellow. The extent of the peroxide breakdown reaction was estimated by the 450nm absorption of the quenched solution[218].

A450 measurements plotted against antibody concentration were fitted to sigmoidal curves (figure R3 8a,b,c,d,e,f). The midpoint of the fitted curve at which the A450 was halfway between baseline and maximum was the estimated EC50 concentration of a given antibody-antigen combination (table R3 4). In theory, if equilibrium is reached during the VHH binding step, a lower EC50 corresponds to a higher affinity but the relationship is not linear and affinity cannot be calculated from EC50. In addition, at higher affinities, EC50 values tend towards a theoretical minimum because at a certain dilution, there will not be enough total antibody to elicit an observable A450 response, therefore affinity differences between already high affinity antibodies can be small enough to be unobservable compared to assay noise.

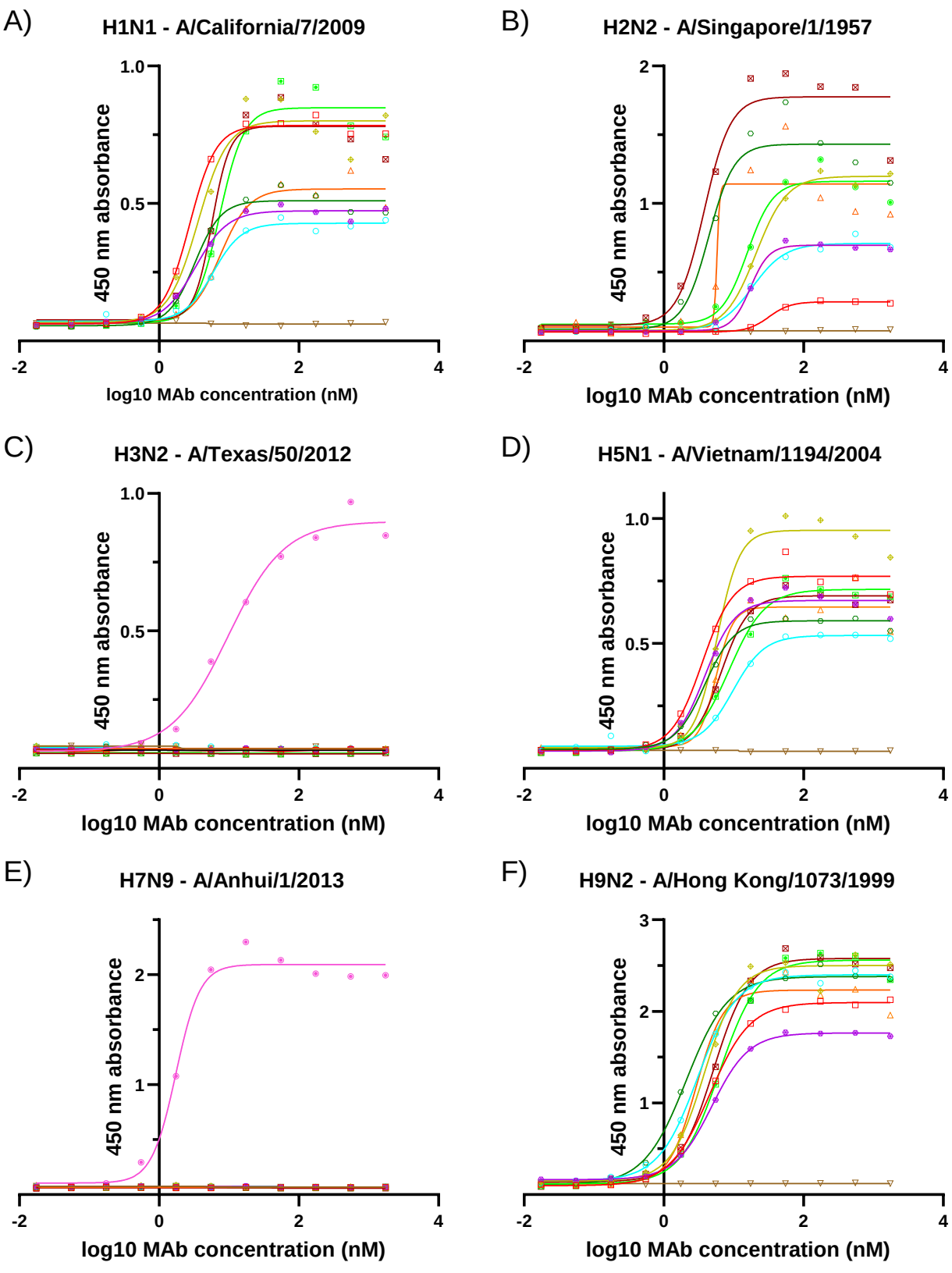

A)
H1N1 - A/California/7/2009
450 nm absorbance
log10 MAb concentration (nM)
B)
H2N2 - A/Singapore/1/1957
450 nm absorbance
log10 MAb concentration (nM)
C)
H3N2 - A/Texas/50/2012
450 nm absorbance
log10 MAb concentration (nM)
D)
H5N1 - A/Vietnam/1194/2004
450 nm absorbance
log10 MAb concentration (nM)
E)
H7N9 - A/Anhui/1/2013
450 nm absorbance
log10 MAb concentration (nM)
F)
H9N2 - A/Hong Kong/1073/1999
450 nm absorbance
log10 MAb concentration (nM)
R1a-B6
VarH (R31I)
VarD (D95N)
VarI (Y102D)
VarJ (Y102S)
VarP (R31I,D95N,Y102D)
VarQ (R31I,D95N,Y102S)
WC21
Positive control
R1a-G2 (negative control)

Figure R3 8a,b,c,d,e,f: ELISA concentration response curves comparing R1a-B6 and variants 96-well plates were coated in antigen standards of **A)** H1N1, **B)** H2N2, **C)** H3N2, **D)** H5N1, **E)** H7N9 and **F)** H9N2 influenza viruses. Plates were then incubated with dilution series of Myc-tagged R1a-B6, R1a-B6 variants and a negative control (R1a-G2) as well as positive controls in the case of H3 (NB3-08)[unpublished data] and H7 (NB7-09)[123]. R1a-B6 shows no detectable binding to H3 or H7. Plates were then incubated with an anti-Myc-tag antibody fused to horseradish peroxidase (HRP). Remaining HRP bound to the plate after washing catalysed a colour change in a substrate which was measured by 450 nm absorbance. A greater 450nm absorbance signal corresponded to a greater amount of VHH bound to HA. Each point on the graphs represents the mean average of 2 wells and a sigmoid curve was fitted to each dilution series. Generally, the further left a curve is, the greater the affinity of VHH for HA. The EC50, the concentration of antibody at which half the saturated antibody response is reached, was used as a measure of antibody binding, and correlates with affinity.

EC50 (nM)	R1a-B6	WC21	VarH	VarD	VarI	VarJ	VarP	VarQ	R1a-G2	Positive control
H1 - A/California/7/2009	2.72	7.44	3.29	3.61	5.94	7.34	3.23	5.69	>1744	N/A
H2 - A/Singapore/1/1957	32.0	5.76	17.3	20.9	18.6	15.6	4.38	3.77	>1744	N/A
H3 - A/Texas/50/2012	>1744	>1744	>1744	>1744	>1744	>1744	>1744	>1744	>1744	9.66
H5 - A/Vietnam/1194/2004	3.52	5.58	3.87	5.74	9.73	8.58	3.83	6.56	>1744	N/A
H7 - A/Anhui/1/2013	>1744	>1744	>1744	>1744	>1744	>1744	>1744	>1744	>1744	1.73
H9 - A/Hong Kong/1073/1999	4.56	2.90	4.86	3.80	3.04	6.27	2.09	5.12	>1744	N/A

Table R3 4: ELISA EC50 values As described in figure R3 8a,b,c,d,e,f, sigmoid curves fitted to mean 450nm absorbance values are used to estimate the concentration of VHH at which the influenza antigen surface is 50% saturated. EC50 values are presented to 3 significant figures. Where the A450 values at the highest concentration of VHH (30µg/ml, 1744nM) was indistinguishable from the A450 value in VHH free wells, the result was recorded as >1744nM. The positive control for H3 (A/Texas/50/2012) antigen was NB3-08 and for H7 (A/Anhui/1/2013) was NB7-09. Positive controls were not used for the other antigens as R1a-B6 binding to these antigens is detectable by ELISA.

R1a-B6 and all variants showed clear binding curves to the H1, H2, H5 and H9 antigens and showed no sign of binding to H3 and H7 antigens at the concentrations tested. R1a-B6 and its variants had relatively similar EC50 values for H1, H5 and H9. Compared to R1a-B6, the EC50 of WC21 was 2.7-fold higher for H1, 1.6-fold higher for H5 and 1.6-fold lower for H9.

Greater ratio differences were seen in the H2 results. EC50 values from H2 antigen ELISAs are plotted in figure R3 9. The 3 antibodies with all 3 key substitutions, WC21, VarP and VarQ all had similar EC50 values, 5.5, 7.3 and 8.5-fold lower than the H2 EC50 of R1a-B6. In between the EC50 of R1a-B6 and the EC50s of WC21, VarP and VarQ were the EC50 values of all of the variants with one key substitution each, VarH, VarD, VarI, VarJ. These results imply that each of the key substitutions alone improves the affinity of R1a-B6 for H2 antigen

and the combined effect of the 3 substitutions is an affinity for H2 similar to that of WC21, providing support for the 3 key substitution hypothesis.

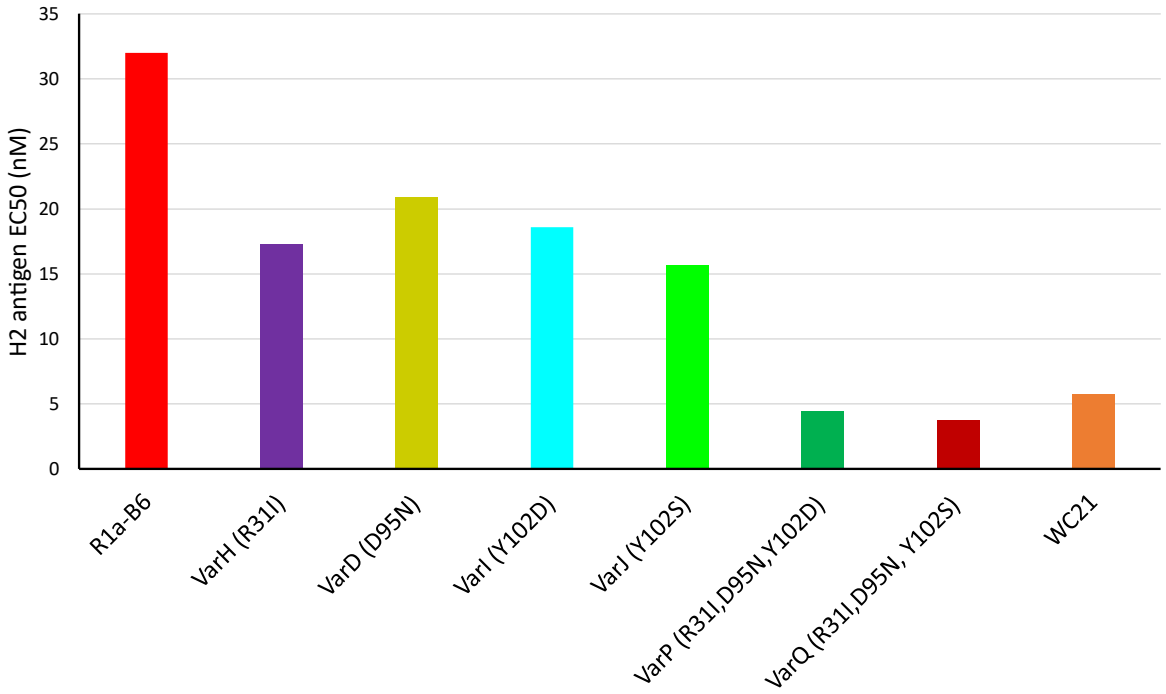


Figure R3 9: Using ELISA data to examine the impact of substitutions in positions 31, 95 and 102 on H2 antigen binding Bar graph showing the estimated EC50 values against H2N2 (A/Singapore/1/1957) antigen of R1a-B6, VarH, VarD, VarI, VarJ, VarP, VarQ, and WC21. R1a-B6 has the highest estimated EC50 of the group indicating the lowest affinity. The VHHs with single substitutions at positions 31, 95 or 102 all show reduced EC50 values compared to R1a-B6. The VHHs with substitutions at all 3 positions, VarP, VarQ and WC21 all have similar EC50 values which are lower than all the single mutant variants.

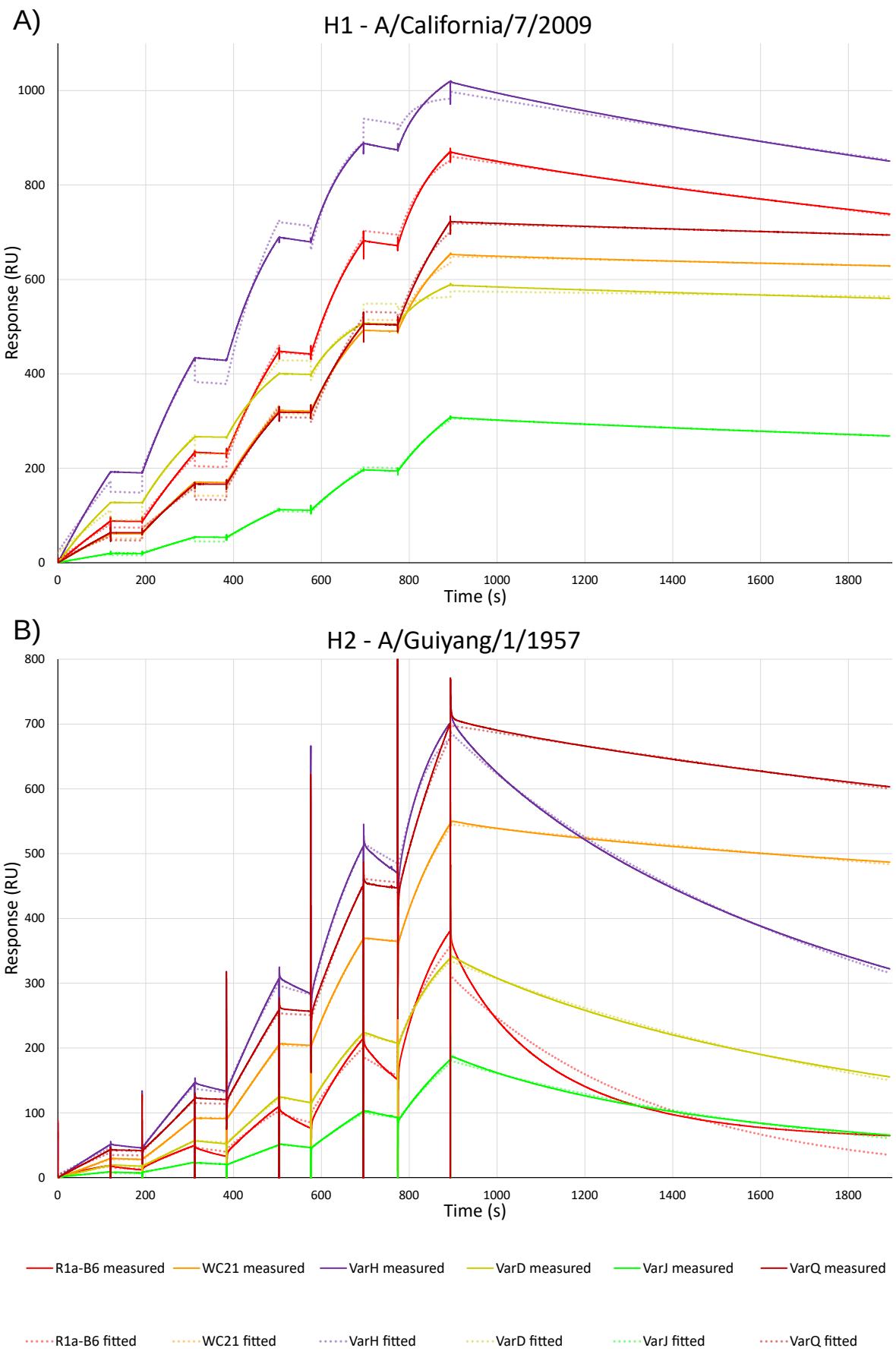
5.7. Investigating the 3 key substitution hypothesis with SPR

To further investigate the 3 hypothesised key substitutions found in WC21, the use of SPR was revisited. Compared to the previous application, in which SPR was used for preliminary screening of R1a-B6 variants from FACS, the number of VHH to be tested was smaller. Only R1a-B6, WC21, VarH, VarD, VarJ and VarQ were tested. Therefore, in these experiments, it was decided to immobilise VHH on SPR chips and use HA as the SPR analyte. The previous set of SPR experiments used VHH as the analyte and HA as the ligand. The benefit of VHH as the ligand is that VHH domains, being much more conformationally stable than HA, could be exposed to regeneration cycles without noticeable damage to the binding surface. Therefore, the ligand surface could be repeatedly used without the previous analyte influencing results.

The binding of recombinant H1, H2, H5 and H9 HAs (A/California/7/2009, A/Guiyang/1/1957, A/Anhui/1/2005 and A/Hong Kong/1073/1999) were tested on immobilised R1a-B6, WC21, VarH, VarD, VarJ and VarQ using an SCK protocol. SPR traces and fitted curves from this set of experiments are plotted in figure R3 10a,b,c,d. Kinetic parameter estimates from the fitted curves are listed in table R3 5a,b,c,d and plotted onto affinity plots in figure R3 11a,b,c,d. Also included in table R3 5a,b,c,d are values of Chi^2 and R_{max}. Chi^2 values are an measure of the deviation of the fitted line response curve from the measured response curve and R_{max} is the estimated maximum response of a saturated surface. If the Chi^2 of a response curve exceeds 10% of R_{max}, the resultant kinetic peramiters should be interpreted with caution.

Immediately apparent in the H2 antigen response curves is the high amplitude peaks and troughs observed when the solution running over the chip is changed. These are seen on the other traces but at much lower magnitude. The H2 antigen used in these experiments, unlike the other antigens, was reconstituted from a lyophilised state. It is possible that additives from the lyophilisation process or undissolved lyophilised antigen contributed to these artifact peaks and troughs. Despite this, the H2 traces clearly show the binding and disassociation of the antigen on the surface and the kinetic differences between the antibodies. Outside of the artifact peaks and troughs, the measured and fitted curves are closely correlated to each other.

Some deviation between fitted and measured response curves is also noticeable for very high affinity interactions, especially those involving H1. Each of the interactions affected by this phenomenon were measured as possessing sub-nanomolar affinity. These interactions were close to the limit of the range of affinities which can be accurately determined by SPR. Given that this phenomenon only affects high affinity interactions, it did not interfere at all with detection of affinity improvements towards H2 and H9 antigens.


A)
H1 - A/California/7/2009
Response (RU)
Time (s)
B)
H2 - A/Guiyang/1/1957
Response (RU)
Time (s)
R1a-B6 measured
WC21 measured
VarH measured
VarD measured
VarJ measured
VarQ measured
R1a-B6 fitted
WC21 fitted
VarH fitted
VarD fitted
VarJ fitted
VarQ fitted

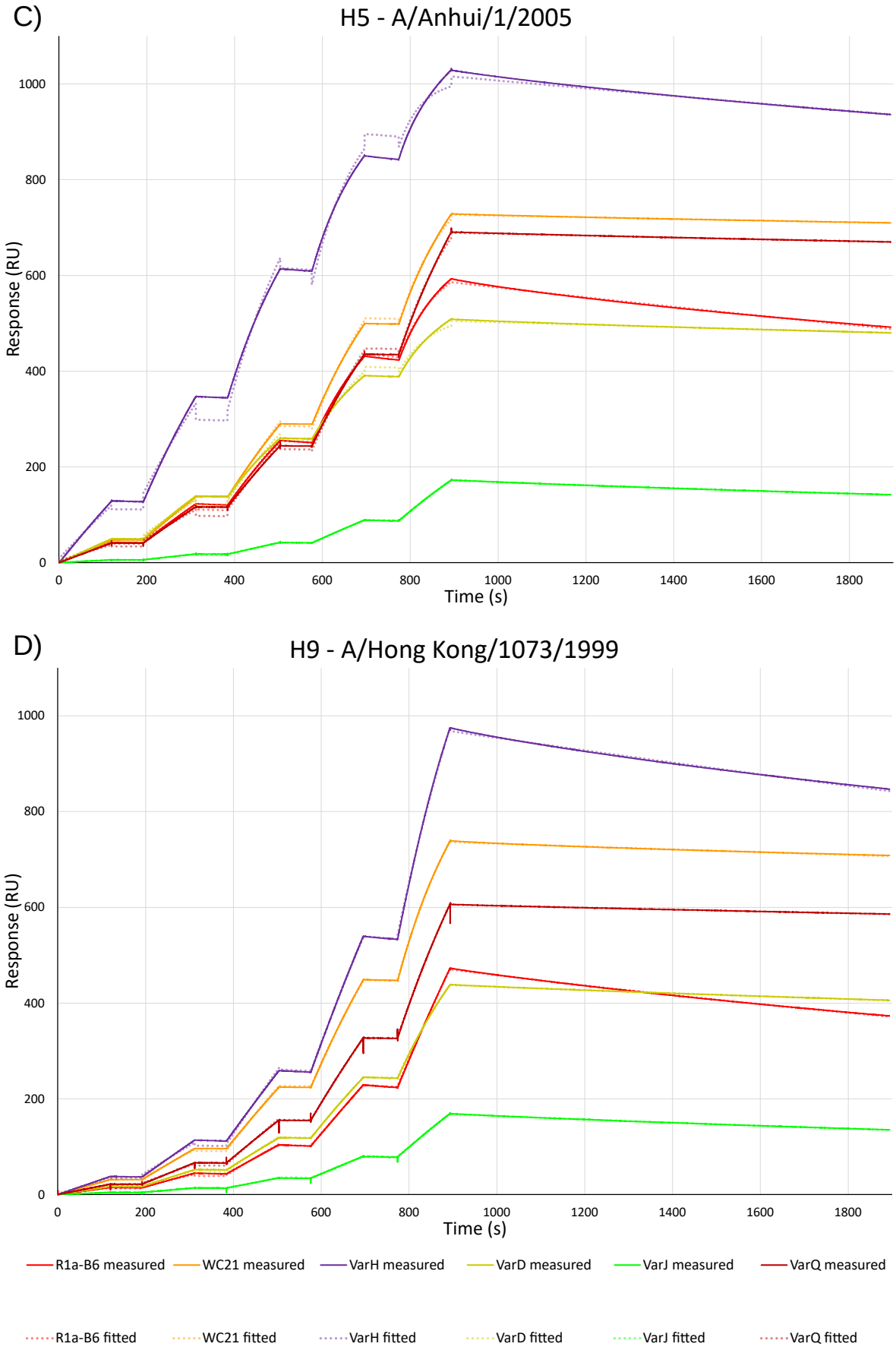


Figure R3 10a,b,c,d: Comparison of HA binding kinetics of R1a-B6 variants using SPR
Traces of SPR chip response, indicative of relative mass of HA bound to immobilised VHH over

time, are used to estimate the kinetic parameters of the interactions of VHH and HA. Solid lines represent the recorded response and dotted lines are software fitted to the recorded data. Each VHH is immobilised on a different chip surface and the quantity of immobilised VHH varies so the absolute magnitude of response is not representative of the kinetics of the interaction. Instead, the shape of the response curves is used to compare the effects of substitutions on the binding kinetics of R1a-B6.

A)

H1 - A/California/7/2009	k_{on} x10^5($M^{-1}s^{-1}$)	k_{off} x10^{-5}(s^{-1})	KD (nM)	R_{max} (RU)	Chi^2 (RU^2)
R1a-B6 (repeat 1)	2.39	16.0	0.669	622	79.1*
R1a-B6 (repeat 2)	2.40	15.0	0.627	822	202*
R1a-B6 (repeat 3)	1.46	15.6	1.07	904	75.3
WC21	1.30	3.11	0.240	687	70.6*
VarH (R31I)	2.73	15.7	0.575	1010	420*
VarD (D95N)	2.85	1.93	0.0677	576	262*
VarJ (Y102S)	0.692	13.2	1.91	393	6.81
VarQ (R31I, D95N, Y102S)	1.02	3.47	0.340	800	101*

B)

H2 - A/Guiyang/1/1957	k_{on} x10^5($M^{-1}s^{-1}$)	k_{off} x10^{-5}(s^{-1})	KD (nM)	R_{max} (RU)	Chi^2 (RU^2)
R1a-B6 (repeat 2)	0.228	220	96.6	648	1120*
WC21	0.405	12.4	3.05	669	38.3
VarH (R31I)	0.924	104	11.2	759	1110*
VarD (D95N)	0.344	80.4	23.4	468	39.5
VarJ (Y102S)	0.141	116	81.9	473	37.1
VarQ (R31I, D95N, Y102S)	0.336	15.2	4.51	909	1080*

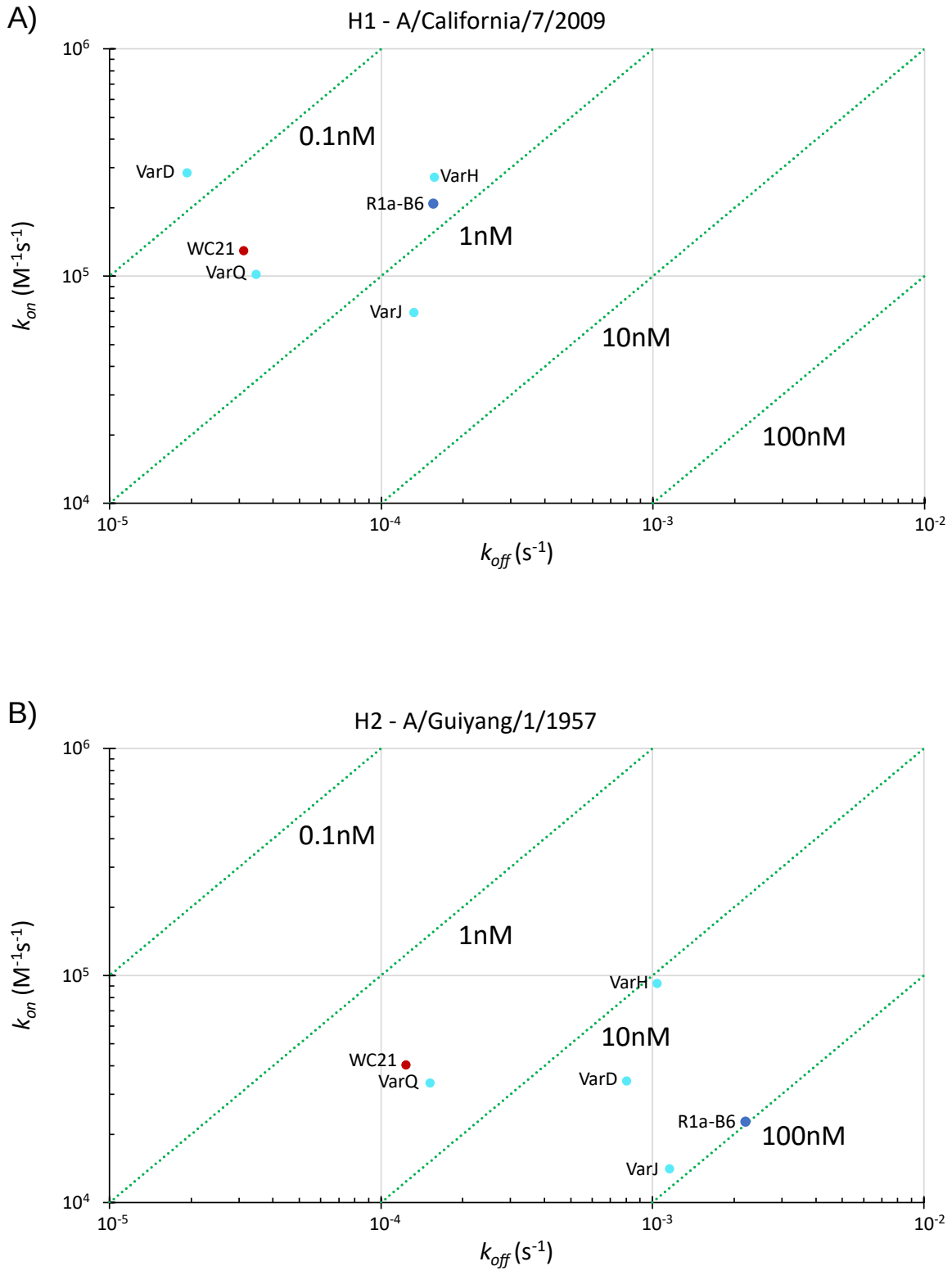
C)

H5 - A/Anhui/1/2005	k_{on} x10^5($M^{-1}s^{-1}$)	k_{off} x10^{-5}(s^{-1})	KD (nM)	R_{max} (RU)	Chi^2 (RU^2)
R1a-B6 (repeat 1)	1.05	18.2	1.73	657	9.89
R1a-B6 (repeat 2)	1.13	17.1	1.51	781	23.7
R1a-B6 (repeat 3)	0.556	16.7	3.00	872	7.13
WC21	0.855	2.38	0.278	850	21.5
VarH (R31I)	1.92	8.13	0.424	1030	241*
VarD (D95N)	1.39	5.01	0.360	529	43.9
VarJ (Y102S)	0.163	19.8	12.2	585	0.264
VarQ (R31I, D95N, Y102S)	0.637	2.85	0.447	904	27.3

D)

H9 - A/Hong Kong/1073/1999	k_{on} x10^5($M^{-1}s^{-1}$)	k_{off} x10^{-5}(s^{-1})	KD (nM)	R_{max} (RU)	Chi^2 (RU^2)
R1a-B6 (repeat 1)	2.10	123	5.85	578	0.497
R1a-B6 (repeat 2)	0.868	108	12.5	1530	1.71
R1a-B6 (repeat 3)	5.23	268	5.13	683	0.773
WC21	0.547	4.21	0.770	1070	2.19
VarH (R31I)	0.359	15.0	4.16	1900	9.68
VarD (D95N)	0.309	7.53	2.47	877	1.46
VarJ (Y102S)	0.114	26.2	23.1	1020	0.377
VarQ (R31I, D95N, Y102S)	0.389	4.65	1.20	1190	3.31

Table R3 5a,b,c,d: Kinetic data from SPR comparing R1a-B6, WC21 and variants containing the 3 key substitutions found in WC21 R1a-B6 repeats 1 and 3 with H2 antigen were omitted due to improbable fitted values. Chi^2 values >10% of the corresponding values of R_{max} are highlighted with an asterisk. All values are recorded to 3 significant figures.


A)
H1 - A/California/7/2009
0.1nM
1nM
10nM
100nM
VarD
VarH
R1a-B6
WC21
VarQ
VarJ
10^6
10^5
10^4
10^{-5}
10^{-4}
10^{-3}
10^{-2}
k_{on} ($M^{-1}s^{-1}$)
k_{off} (s^{-1})
B)
H2 - A/Guiyang/1/1957
0.1nM
1nM
10nM
100nM
VarH
WC21
VarQ
VarD
R1a-B6
VarJ
10^6
10^5
10^4
10^{-5}
10^{-4}
10^{-3}
10^{-2}
k_{on} ($M^{-1}s^{-1}$)
k_{off} (s^{-1})

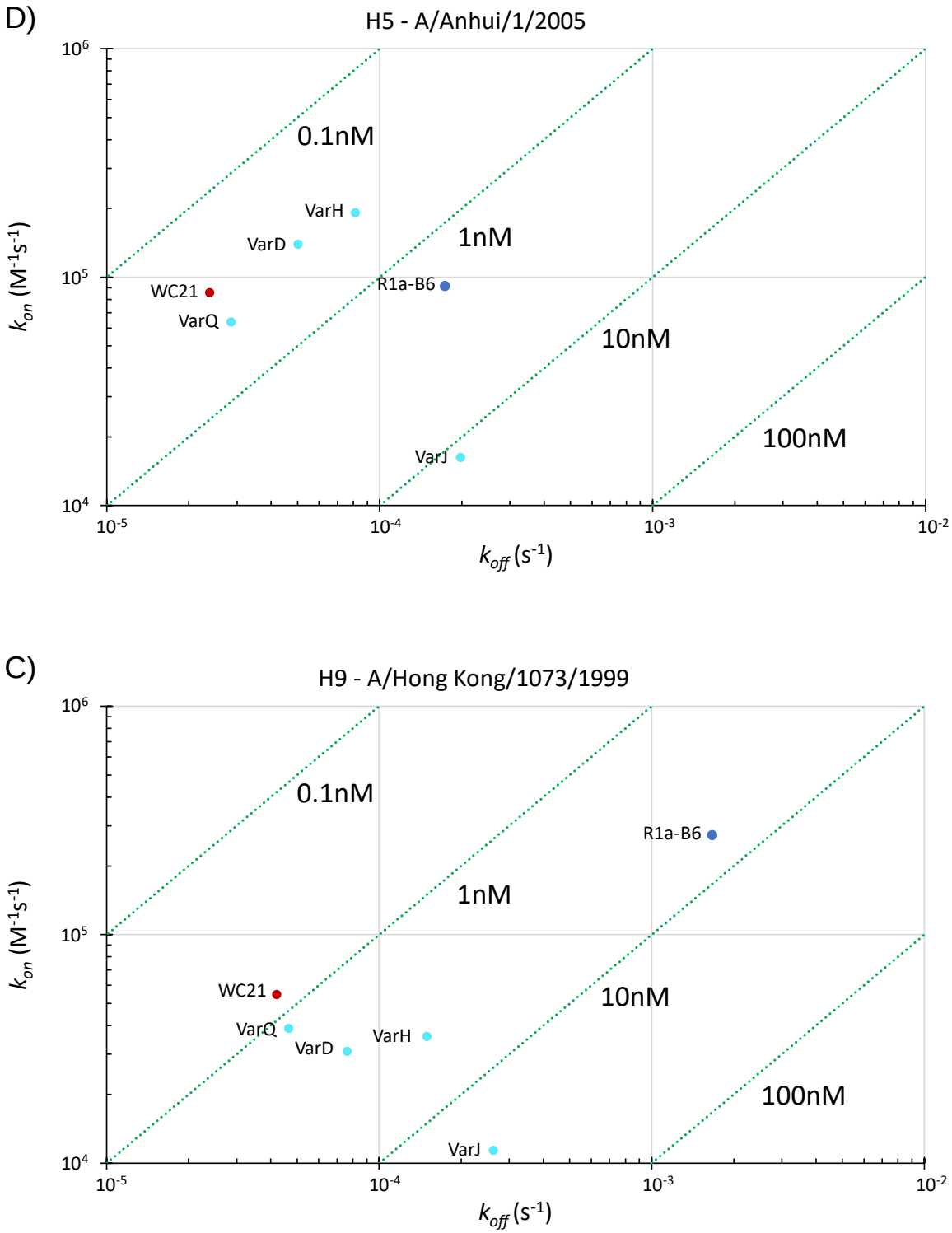


Figure R3 11a,b,c,d: Use of SPR to test the 3 substitution hypothesis and to parse the role of each substitution in change in K_{on} and K_{off} Purified VHH was immobilised to the surface of SPR chips. Using these SPR chips, binding kinetics to recombinant **A)** H1, **B)** H2, **C)** H5 and **D)** H9 HA was measured. As with figure R3 3a,b,c, isoaffinity diagonals are plotted across the graph and marked with the affinities they represent.

As was seen in previous SPR results, WC21 showed affinity improvements for all subtypes tested, predominantly driven by a slower k_{off}. In agreement with H2 PV titre and ELISA EC50 reductions, WC21 has a dramatically higher affinity (>30-fold improvement) for H2 antigen than R1a-B6 does.

In strong support for the 3 key substitution hypothesis, WC21 k_{off} and k_{on} for all 4 antigens are tightly correlated with those of VarQ (R31I, D95N, Y102S), the R1a-B6 triple mutant, suggesting that the extra substitutions on WC21 not possessed by VarQ (G26E and S56T) have little or no impact on the binding kinetics of WC21.

The 3 single mutant variants, VarH (R31I), VarD (D95N) and VarJ (Y102S) were included in the experiment to determine their individual contributions to group 1 antigen binding kinetics. From this data it seems that R31I and D95N in isolation show the greatest improvements to R1a-B6 affinity across group 1 antigens. Across all subtypes tested, to varying degrees, these two substitutions were linked with an increase in affinity. Interestingly, for H1 antigen, the D95N substitution alone (VarD) improved affinity >10-fold while WC21/VarQ with all 3 key substitutions only achieved 2 to 3-fold increase.

Y102S, like R31I and D95N slows R1a-B6 dissociation from H2 and H9 antigen but, unlike the other two, has little effect on H1 or H5 dissociation. In addition, Y102S was measured to decrease k_{on} for all 4 antigens, offsetting the affinity improvements to H2 and H9. This data contrasts with the H2 antigen ELISA EC50 values of these same variants which imply that Y102S improves H2 binding to a similar extent as R31I and D95N.

The effect of single substitutions may not be the same as those same substitutions in combination if there is an epistatic interaction between them[210]. So it may be that Y102S contributes more substantially to the binding breadth of WC21 and VarQ when accompanied by R31I and D95N. The enrichment of this substitution in the H9 sorted library H9R3b combined with the ELISA data and the k_{off} improvement observed for H2 and H9 antigens all lend credence to the hypothesis that Y102S contributes to the functional improvement of WC21.

6. Results chapter 4: Epitope footprint mapping and antigenic differences between subtypes

6.1. Chapter introduction

The data from the previous chapter demonstrated that the R1a-B6 variants WC11 and WC21 possess a markedly improved ability to bind HA, particularly H9 and H2 subtypes. These affinity improvements were shown to translate into far greater improvement in H9 and H2 PV neutralisation potency whilst maintaining potency against H1 and H5 viruses. Furthermore, these improvements were demonstrated to be in major part attributable to the substitutions R31I, D95N and Y102D/S. However, measurement of affinity improvement and localising the substitutions responsible gives little insight into the structural mechanisms of the broadening of the antibody reactivity range. The experiments in this chapter aimed to identify the sequence and structural features which distinguish the stem epitopes of H1, H2, H5 and H9 HA, to identify how the barriers to high affinity H9 and H2 binding are overcome by the affinity matured R1a-B6 variants WC11 and WC21.

This section follows up on deep mutational scanning experiments carried out in[102]. In the study, a library of error prone PCR mutagenised HA was displayed on yeast. Yeast clones unable to bind to VHH were sorted by FACS and sequenced by Illumina sequencing. Two of the VHH tested in this study were R1a-B6 and clonally related antibody R2b-E8. Mutations identified as highly enriched by the sorting steps between the unsorted library and the sorted libraries were assessed to be likely to interfere with the antibody-antigen interaction. This approach to epitope mapping is commonly used in the literature[219-221].

The F45 residue present in H2 HA (H1 residue - I45) was identified as a key barrier to stem epitope VHH binding. Of the panel of 5 stem binding VHH (including R1a-B6 and R2b-E8), only R2b-D9 could bind to yeast displaying H1 (A/California/7/2009) with the I45F substitution. Furthermore, the D46N polymorphism, present in some H1 isolates, was found to abrogate R2b-E8 binding but not R1a-B6, explaining some of the observed differences in H1 binding range between R2b-E8 and R1a-B6.

A similar experiment was carried out in the present study, focussing on the R1a-B6 clonal lineage, with R1a-B6, R2b-E8, WC11 and WC21. In this experiment however, a saturated mutagenesis library was used, in theory allowing the effect of every single amino acid polymorphism in the epitope area to be interrogated. The aim was to map in detail the residues involved in binding of each antibody to HA and to analyse the differences uncovered. In addition, to identify mutations which have a differential effect on binding to the different related antibodies and to attempt to relate that to sequence variation between HA subtypes.

6.2. Deep mutational scanning

6.2.1. Library creation and experimental design

Like in [102], the EBY100 strain of yeast was used to display a library of mutated H1 A/California/7/2009 HA0 variants which were, as with the VHH, encoded on the pNIBS-5 plasmid. This library was stained for display of HA and binding of VHH to expressed HA (see figure R4 1a,b table MET 7b and figure MET 3). FACS was used to isolate yeast which expressed HA to a high level (strong AF488 staining) but did not show signs of VHH binding (weak/no AF647 staining) (see figure R4 1c). Thereby, it was hoped that the sorted yeast encoded HA gene mutations which disrupted binding to the VHH they were stained with.

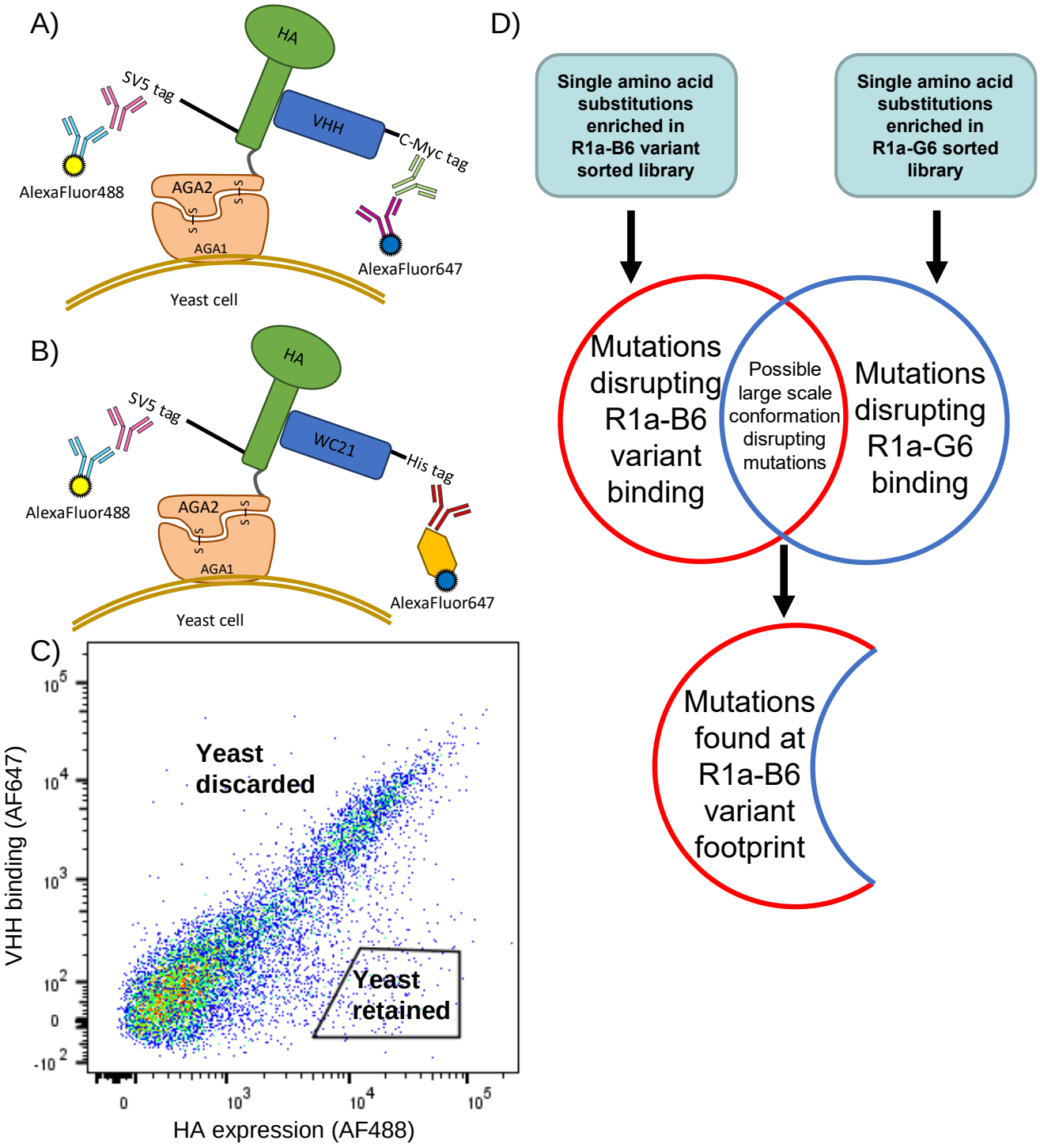


Figure R4 1a,b,c,d: Use of yeast display and FACS for epitope mapping of R1a-B6 and R1a-B6 variants Yeast displaying a saturated mutagenesis library of H1 HA (A/California/7/2009) were stained with 100nM VHH, secondary antibodies and fluorophores. The AF488 fluorescence of the stained population corresponded to the level of HA display. The level of AF647 fluorescence corresponded to the amount of VHH bound to the expressed HA. **A)** Cartoon representation of yeast stained with flow cytometry staining protocol 3a (see table MET 7b and figure MET 3). This protocol was used to stain the mutagenised library with R1a-B6, R2b-E8, WC11 and R1a-G6 (a HA head binding antibody unrelated to R1a-B6). **B)** Cartoon representation of yeast stained with flow cytometry staining protocol 3b. WC21 batch 1, which was used for this experiment, was missing the Myc-tag, necessitating a protocol with a His-tag binding secondary

antibody. The two protocols were functionally equivalent. **C)** Example dot plot showing the sorting process. The yeast retained from the sort were positive for HA display but negative for VHH binding. Therefore, the sorts enriched yeast displaying HA with mutations at residues important to the binding process in the epitope footprint of the staining VHH. The same mutagenised HA library was sorted using this method while stained with R1a-B6, R2b-E8, WC11, WC21 and R1a-G6. **D)** Illumina sequencing of the unsorted library and the sorted libraries was used to identify single amino acid substitutions which were enriched in each sorted population and therefore the mutations which disrupted the binding of the corresponding antibodies. The experiment aimed to map the key residues interacting with R1a-B6 and related antibodies in the HA stem. R1a-G6 was included to identify mutations disruptive to the HA structure as a whole. The R1a-G6 epitope is distant from the stem so any mutations enriched in both the R1a-G6 sorted library and one of the stem binder sorted libraries may be a mutation causing wide ranging structural disruption.

The method of library construction, however, was different. This project made use of several saturated mutagenesis sub-libraries based on A/California/7/2009 produced by Tiziano Gaiotto[unpublished data]. Each sub-library was mutagenised with the aim of introducing each of the 19 non-wild type amino acids at every position over a span of approximately 50 residues using mutagenic primers (similar in principle to the site directed mutagenesis used to introduce alanine substitutions described in section 2.2.1). Mutagenised pNIBS-5 plasmid was transformed first into *E.coli* and then into EBY100 yeast. Transformed yeast were grown to stationary phase and frozen at -80°C with glycerol as described in section 2.1.5. Sub-libraries were produced to mutagenise every amino acid from HA1 D11 to HA2 Y184 in 11 approximately 50 amino acid spans. By pooling the appropriate sub-libraries, the effects of mutations in specific parts of HA can be examined while minimising the inclusion of mutations in off-target parts of the molecule. Saturation mutagenised libraries, in theory, allows interrogation of the effect of every single amino acid substitution to the mutagenised regions of a protein[222]. By contrast, error prone PCR yields only a limited set of mutations at each codon so sequence coverage is less thorough[223]. Furthermore, the entire epPCR amplicon is mutagenised, often resulting in a less targeted mutagenised area.

The decision of which sub libraries to combine and sort was based both on experimental data collected from R1a-B6 and R2b-E8 and structural data from similar stem binding antibodies. The experiments in [102] demonstrated that mutations at positions HA2 20, 21 and 45 were able to disrupt R1a-B6 binding while R2b-E8 binding was disrupted by mutations at HA2 positions 20, 21, 45, 46, 48, 52, 53, so it was essential that the mutagenised regions of the library

covered all of these residues. In addition, x-ray crystal structures of other group 1 subtype cross neutralising stem binding antibodies were used to locate other HA residues which were likely to make contact with R1a-B6 and related VHH. Structures of CR6261, C179 and F10 in complex with HA (PDB codes 3GBN[67], 4HLZ[66], 3FKU[69]) were analysed in the online tool PDBePISA[143, 224]. Residues on HA which had solvent accessible surface area occluded by 1 or more of these antibodies were deemed to be important to mutagenise as part of the deep mutational scanning experiment. These residues were as follows HA1 18, 38, 39, 40, 41, 42, 291, 292, 293, 318, 319 and HA2 17, 18, 19, 20, 21, 36, 37, 38, 39, 41, 42, 43, 44, 45, 46, 48, 49, 50, 52, 53, 56. These residues are mapped onto the structure of A/California/7/2009 in figure R4 2a.

Based on the criteria of including all of these residues in the pooled library mutagenised region, the sub-libraries spanning residues HA1 11 to 56, HA1 239 to 284, HA1 285 to HA2 2, HA2 3 to 49 and HA2 50 to 96 were pooled by inoculating 2ml of SD/CAA with 10µl of each of the frozen glycerol stocks of the sub-libraries. This inoculated medium was grown to stationary phase and refrozen with glycerol to create a pooled library stock. The residues mutagenised in the pooled saturated mutagenesis library are shown in figure R4 2b.

One of the limitations of next generation sequencing is the maximum read length[225]. For this project the maximum read length that was available was 300bp. As with the VHH sequencing, reads started from both ends of the sequencing amplicon, allowing merging of forward and reverse reads, extending the maximum length of a readable amplicon beyond 300bp. Despite this, it was not practical or useful to sequence the whole HA molecule. Instead, two regions were chosen, referred to as amplicon A and amplicon B. These two amplicons covered the residues listed above which were occluded or partially occluded by binding of CR6261, C179 and/or F10. HA amplicon A spans HA1 Y17 to L42 and HA amplicon B spans HA1 S291 to HA2 I56. Amplicons A and B are mapped onto the A/California/7/2009 HA structure in figure R4 2c.

The protocol used in [102] involving 3 rounds of FACS selection was condensed into a single round. It was thought that with more than one round of selection, false positive enrichments may occur due to passenger mutations being enriched

alongside mutations which abrogate binding[225]. With a single round of selection, the maximum enrichment factor of a single clone is much lower, potentially mitigating this issue. Furthermore, a single round of selection results in a quicker experiment.

In addition to the 4 R1a-B6 clonal lineage VHH antibodies, the mutagenised HA library was also sorted for binding defects to a head binding VHH, R1a-G6. R1a-G6 binds to a site distant to R1a-B6 (see figure R4 2b) so no residue is likely to directly interact with both R1a-G6 and a stem binding VHH. Therefore, overlap in residues enriched in mutations between the R1a-G6 sorted library and the stem binder sorted libraries are likely to be mutations causing long range structural distortion to the HA molecule. As the deep mutational scan aims to identify residues directly interacting with stem binding antibodies, residues enriched in mutations in the R1a-G6 sorted library were excluded from further analysis (see figure R4 1d). In the previous study, R1a-G6 was used for the same purpose but the implementation was different. Instead of filtering out HA variants disrupting R1a-G6 binding in data analysis, R1a-G6 binding of HA variants was verified in a final sorting step after two rounds of sorting with stem binding VHH.

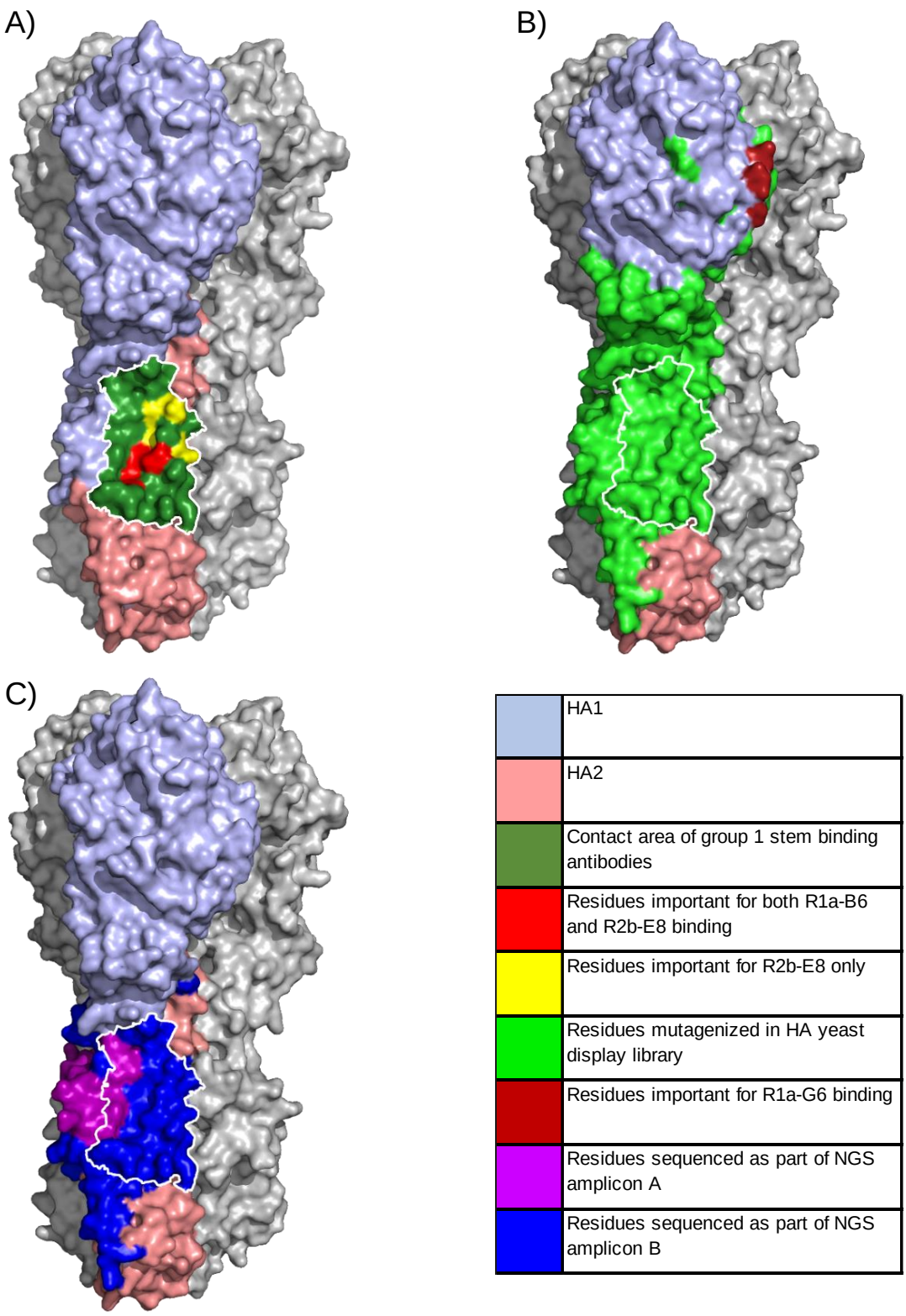


Figure R4 2a,b,c: Structure of HA (A/California/7/2009) annotated with regions important to epitope mapping of the R1a-B6 clonal lineage A) shows both the confirmed and potential contact footprints of R1a-B6 and R2b-E8 mapped onto HA (PDB code 3LZG[29]). Previous epitope mapping data from [102] identified the residues HA1 glycine 20, tryptophan 21 and isoleucine 45 (in red on the figure) as important to R1a-B6 binding and the same 3 plus an additional 4 residues (yellow) were identified as important to R2b-E8 binding. Antibodies similar in epitope and neutralisation range are found in the literature. Those with published structures in complex with HA were used to predict other residues potentially important to R1a-B6 clonal lineage antibody

binding. The residues of HA which were fully or partially occluded by antibodies CR6261, C179 or F10 in their respective antibody-HA complex structures (PDB codes 3GBN, 4HLZ, 3FKU) are highlighted in dark green (analysis of complex structures done using the PBDePISA online tool[143, 224]). These residues are also outlined in white in **A)**, **B)** and **C)**. **B)** shows the regions of the HA molecule covered by the saturated mutagenesis library used to pan for epitope footprint mutations (bright green). All the predicted epitope footprint residues (white outline) were mutagenised in the library used for sorting. Residues previously identified as important for R1a-G6 binding[102] are coloured maroon and do not overlap with the mutagenised residues. **C)** shows the two regions covered by the two Illumina sequencing amplicons used to analyse the results of the epitope mapping sorting experiments. Again, all residues in the white outlined predicted epitope footprint region were analysed as part of HA amplicon A (purple) or B (dark blue), allowing examination of the enrichment or depletion of mutations at each of these positions.

6.2.2. Sorting of the mutagenised HA library

The previously prepared pooled mutagenised HA library was grown up and surface protein display was induced along with yeast displaying wild type A/California/7/2009. 2 million cells from the mutagenised library and 400,000 cells displaying wild type (WT) HA were stained with each VHH according to the protocols in section 2.4 (protocol scaled down accordingly for the WT HA displaying yeast). Staining with R1a-B6, R2b-E8, WC11 and R1a-G6 was done using protocol 3a. Due to the loss of the Myc-tag from WC21 batch 1, the similar and functionally equivalent protocol 3b was used (see figure R4 1a,b table MET 7b and figure MET 3).

Stained cells displaying WT HA were used to set the sorting gate for sorting of the mutagenised libraries. The sort aimed to heavily deplete the mutagenised libraries of HA genes with WT or near WT affinities to VHH so it was important to set the gate such that almost none of the WT HA displaying yeast would land inside the gate in any of the stained populations (see figure R4 3). Following gate setting, the stained mutagenised libraries were all sorted using the same gate. Due to an error in the sorting process, the majority of the R1a-B6 and WC11 stained libraries were sorted incorrectly so the number of correctly sorted cells is much lower compared to the other stained libraries. Gating is shown in figure R4 3 and numbers of cells sorted is recorded in table R4 1.

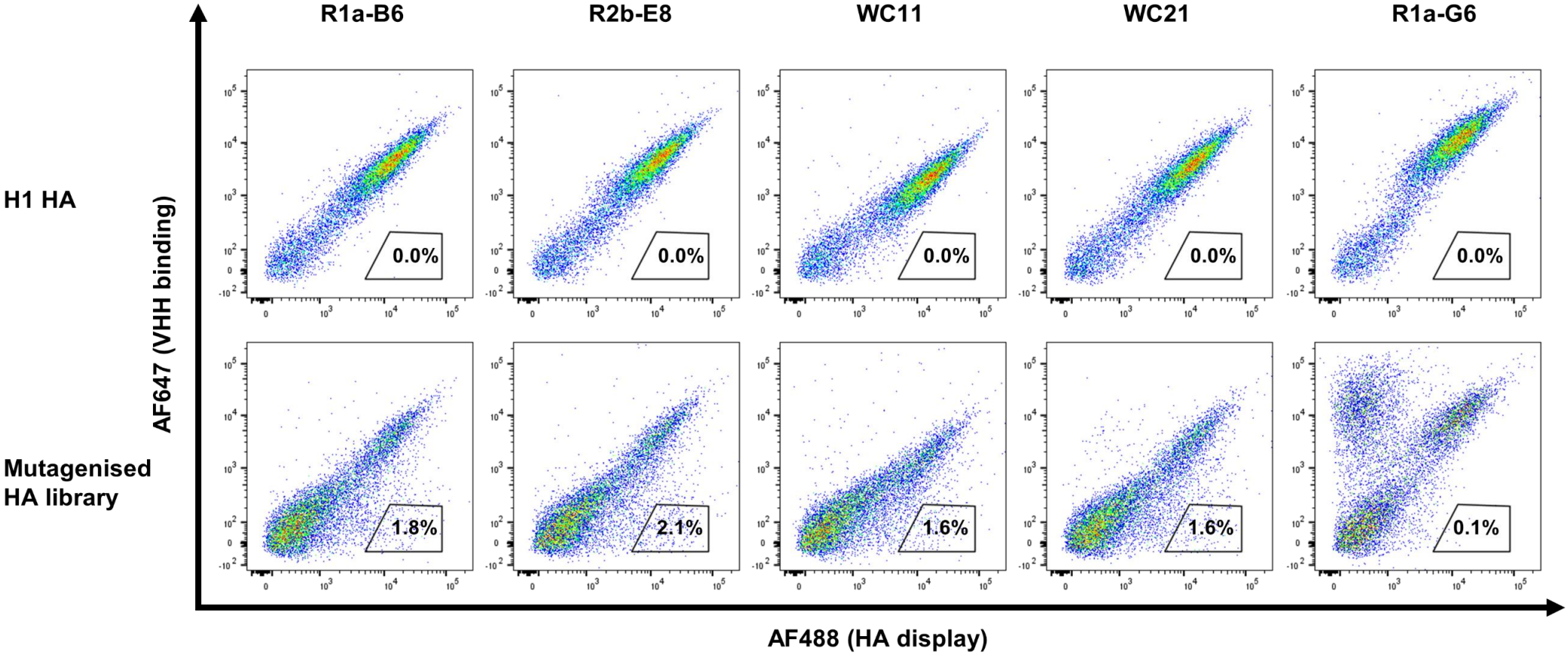


Figure R4 3: Sorting VHH non-binders from a library of mutagenised HA Dot plots show populations of yeast expressing H1 HA (A/California/7/2009) without (top) and with (bottom) mutagenesis. Yeast were stained for HA display and VHH binding to the expressed HA. Each dot plot shows a sample of 10,000 measured events. The gate used to sort the mutagenised stained populations is marked on each plot labelled with the percentage of events inside the gate (to 1d.p.). The gate was positioned so that nearly zero of the yeast expressing WT HA would be found in the gate. Therefore, in the mutagenised HA library, only the yeast displaying HA that was unable to bind VHH were found in the gate and retained in the sorted populations.

	R1a-B6	R2b-E8	WC11	WC21	R1a-G6
Cells sorted	69,000	980,000	63,000	800,000	820,000
Cells retained after sorting	1300	21,000	1000	13,000	610
Percentage of cells retained	1.8	2.1	1.6	1.6	0.1

Table R4 1: Epitope mapping sorting of a mutagenised HA library stained with various VHH antibodies The number of sorted cells, the number of cells retained in the sorted sub-library and the percentage of cells retained after sorting is listed. The number of cells sorted of the R1a-B6 and WC11 sorts is much smaller as the sort had to be repeated with a reset gate.

It is notable that the proportion of cells sorted in the gate differs depending on the VHH antibody used to stain the cells. Only 0.1% of the R1a-G6 stained cells were found in the sorting gate. This is likely attributable to little or none of the R1a-G6 epitope being part of the mutagenised region of HA, meaning the mutations disruptive to R1a-G6 binding would be infrequent in this library. In line with the observation that several mutations (e.g. HA2 D46N, I48S and N53K) have previously been characterised to disrupt R2b-E8 binding but not R1a-B6 binding, a higher proportion of the library fell inside the gate when stained with R2b-E8 than when stained with R1a-B6[102]. Similarly, the gates for WC11 and WC21 enclosed a smaller proportion of the library than either naturally affinity matured antibody, suggesting that the range of disruptive mutations to the WC11 and WC21 binding interactions may be narrower still.

As described in section 2.4.5, sorted cells from the stained library were grown to stationary phase for 48 hours in 10ml SD/CAA medium and frozen at -80^{o}C with glycerol.

Each sorted library, the unsorted library and a sample of yeast encoding WT A/California/7/2009 were sequenced by Illumina sequencing across amplicons A and B. The process of extracting plasmids, amplifying sequences and preparing samples for Illumina sequencing is described in section 2.5.

6.2.3. Processing and analysis of sequence data

The output sequence data was trimmed to remove adapter sequences, forward and reverse reads were merged and unpaired reads were discarded. Merged reads were translated into protein sequence data.

For this analysis, only full-length sequences containing a single amino acid substitution were counted towards enrichment factors. Those with multiple mutations or any insertions or deletions were not counted. The combined sequence noise of introduced mutations during yeast growth, PCR amplification and sequencing was accounted for by using the sequence data from the WT A/California/7/2009 sample (henceforth referred to as the H1 control) which would have been homogeneous before these sources of sequence noise. Enrichment factors were calculated by dividing the post-sort frequency of a mutation by the pre-sort frequency. These techniques were applied both to calculating enrichment factors of mutations at a given position (e.g. combined enrichment of all mutations at position HA1 40) and to calculating enrichment of one specific substitution (e.g. enrichment of mutation HA1 V40K).

Positions in amplicons A and B which were found to be enriched with mutations in the R1a-G6 sorted library (HA1 297, 321, HA2 4, 40 and 55) were excluded from further analysis. Enrichment factors of mutations in R1a-B6, R2b-E8, WC11 and WC21 sorted libraries for all other residues within HA amplicons A and B are shown in table R4 2a,b,c,d. Enrichment factors in this table represent total enrichment of mutations at each residue position (e.g., HA1 V40 (WT residue) -> HA1 X40). Enrichment hotspots in this study were defined as those where total mutations at a position were enriched 2-fold or greater for any of the VHH tested (R1a-B6, R2b-E8, WC11, WC21). Given that fewer than 0.1% of yeast displaying HA sequences with wild type affinity for VHH would have landed in the gate (see figure R4 3) during the sorting process, a 2-fold enrichment in the sorted library was considered evidence that mutations at this position caused a substantial binding defect.

A)

Residue number	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42
WT amino acid	Y	H	A	N	N	S	T	D	T	V	D	T	V	L	E	K	N	V	T	V	T	H	S	V	N	L
R1a-B6	0.1	0.6	0.2	0.0	0.0	0.0	0.0	1.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.8	3.4	2.9	0.0	0.0	0.0	0.0
R2b-E8	0.0	0.5	1.0	0.1	0.2	0.0	0.1	0.4	0.0	1.9	0.1	0.0	0.0	0.0	0.0	0.0	0.1	0.2	0.2	0.8	3.4	7.5	0.1	2.1	0.0	1.0
WC11	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	4.3	0.0	6.9	0.0	0.0	0.0	0.1	0.0	0.1	0.0	0.0	0.0	0.1	0.0	0.8	0.0	0.2
WC21	0.0	0.4	0.0	0.2	0.5	0.0	0.0	0.6	0.1	2.3	0.0	0.2	0.1	0.0	0.0	0.0	0.0	0.2	0.0	1.3	2.1	1.3	0.0	0.0	0.0	0.1

B)

Residue number	291	292	293	294	295	296	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	322	323	324	325	326	327	328	329
WT amino acid	S	L	P	F	Q	N	H	P	I	T	I	G	K	C	P	K	Y	V	K	S	T	K	L	R	L	A	T	G	L	N	I	P	S	I	Q	S	R
R1a-B6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.1	0.0	0.0	0.0	0.0	0.0	0.0	0.7	0.0	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.8	3.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
R2b-E8	0.0	0.0	0.0	0.2	0.3	0.1	0.1	0.4	0.0	0.0	0.7	0.2	0.1	0.3	0.3	0.1	0.3	0.2	0.0	0.1	0.0	0.1	0.3	0.0	1.4	1.5	1.6	2.6	0.4	0.0	0.0	0.1	0.0	0.0	0.1	0.0	0.0
WC11	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.4	2.9	1.4	0.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
WC21	0.0	0.0	0.3	0.2	0.3	0.1	0.0	0.3	0.0	0.0	0.1	0.2	0.0	0.3	0.2	0.0	1.1	0.0	0.0	0.0	0.0	0.0	0.3	0.0	1.8	1.2	0.3	3.7	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

C)

Residue number	1	2	3	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
WT amino acid	G	L	F	A	I	A	G	F	I	E	G	G	W	T	G	M	V	D	G	W	Y	G	Y	H	H	Q	N
R1a-B6	0.1	0.2	0.0	0.2	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.2	0.0	0.0	12.6	7.2	2.1	5.5	0.0	0.0	0.0	0.0	0.0
R2b-E8	0.0	0.1	0.1	0.1	0.1	0.1	0.0	0.0	0.1	0.0	0.1	0.1	0.0	0.0	0.0	0.2	0.0	0.0	12.2	12.0	4.3	8.9	0.8	0.0	0.0	0.0	0.1
WC11	0.0	0.1	0.1	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.0	0.0	0.0	0.0	9.2	8.9	1.8	6.0	0.1	0.0	0.0	0.0	0.3
WC21	0.0	0.1	0.0	0.3	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.0	0.0	0.1	0.0	0.0	0.0	14.6	4.0	0.9	3.7	0.9	0.2	0.0	0.0	0.4

D)

Residue number	29	30	31	32	33	34	35	36	37	38	39	41	42	43	44	45	46	47	48	49	50	51	52	53	54	56
WT amino acid	E	Q	G	S	G	Y	A	A	D	L	K	T	Q	N	A	I	D	E	I	T	N	K	V	N	S	I
R1a-B6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.2	0.0	0.0	0.0	6.3	11.5	0.0	10.9	13.2	1.2	0.4	1.1	5.6	0.0	0.1	1.1	0.0	0.0	0.0
R2b-E8	0.0	0.1	0.0	0.0	0.0	0.0	0.0	2.7	1.0	0.0	0.0	8.4	12.3	0.1	10.6	9.3	3.1	0.3	4.6	6.5	0.0	0.7	1.1	4.4	0.2	0.0
WC11	0.0	0.0	0.6	0.0	0.0	0.8	0.0	1.1	0.7	0.0	0.0	9.4	18.8	0.1	2.6	9.7	1.4	0.0	4.6	6.3	0.0	0.3	0.5	0.0	0.0	0.0
WC21	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.1	0.0	8.0	13.9	0.0	3.1	15.2	1.2	0.2	1.9	6.3	0.0	0.0	0.0	0.0	0.0	0.1

Table R4 2a,b,c,d: Enrichment of mutations between the unsorted HA library and the sorted libraries For each Illumina sequencing dataset, sequences encoding a single amino acid difference from the wild type were recorded and sorted by residue changed. After adjusting for mutations introduced into the sequences by the PCR and sequencing processes, the fold change in mutational frequency at each amino acid position between the unsorted library and each sorted library was calculated to give the enrichment factor displayed in these tables. Mutations at positions with no influence on binding were depleted while mutations disrupting binding were enriched. Cells are colour-coded according to enrichment factor (<1 – green, 1 to 2 – yellow, 2 to 5 – light orange, 5 to 10 – dark orange, >10 – red). **A)** shows the HA1 residues in HA amplicon A. **B)** shows HA1 residues from HA amplicon B. **C)** and **D)** show HA2 residues in HA amplicon B. Enrichment factors are recorded to 1 decimal place. Mutations at positions HA1 297, 321, HA2 4, 40 and 55 were enriched in the R1a-G6 sorted library, therefore, they were omitted from further analysis and from this table.

Details of how this Illumina sequencing data was processed can be found in section 2.5.2. and the Illumina sequencing data pipeline used for this data is summarised in figure R4 4.

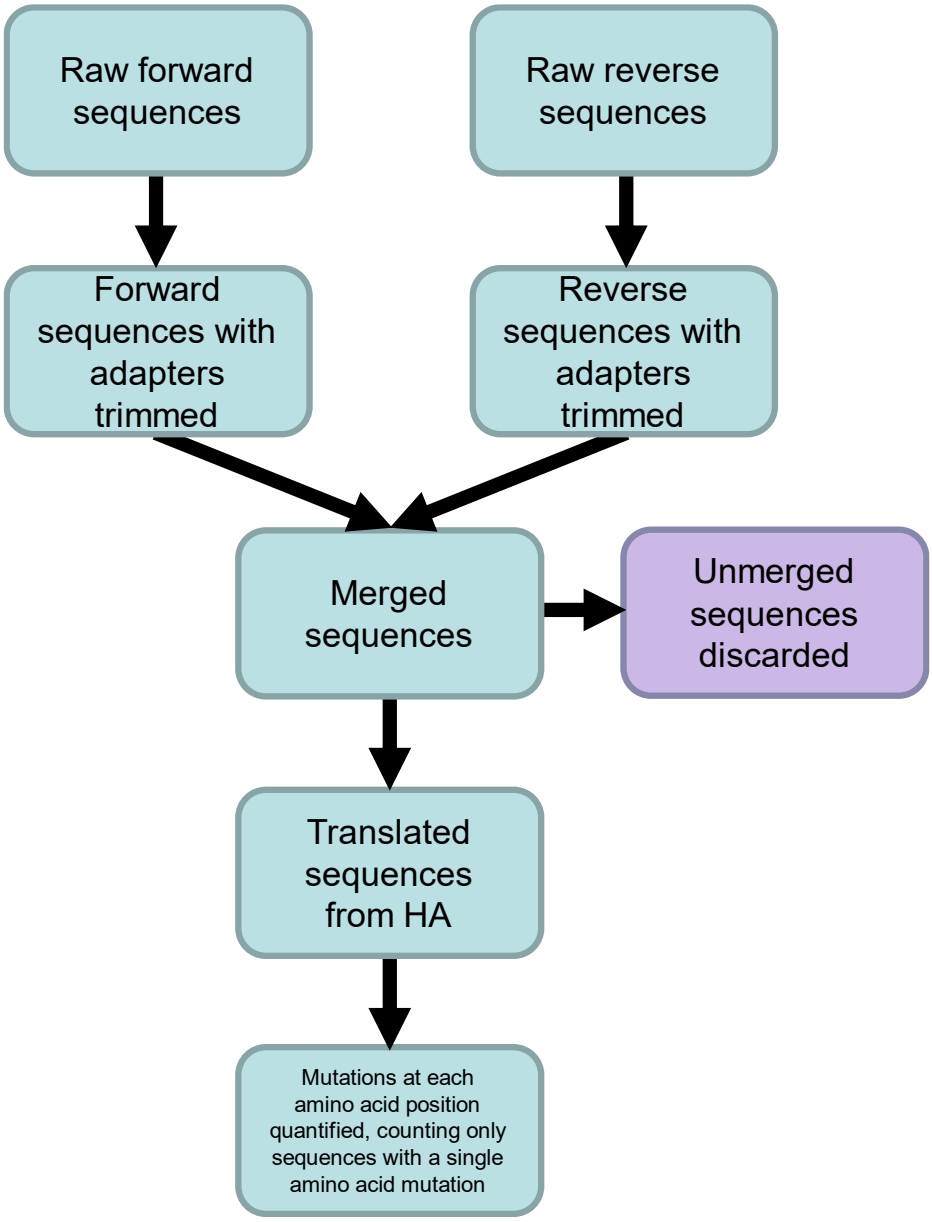


Figure R4 4: Flow chart of the pipeline used to process HA epitope mapping Illumina sequencing data This flow chart applies to the processing of data from amplicon A and amplicon B. As with the VHH library datasets, forward and reverse sequences were trimmed and merged. Processing from this point differs slightly as only sequences with a single substitution mutation were counted towards mutation frequencies in the epitope mapping experiment.

Despite a different library, with a different balance of mutations at each position, and different sorting strategies, there is considerable agreement between the results of the deep mutational scan in [102] and the experiment in this project. Except for positions HA2 44 and 49, the most enriched positions in the sorted

R1a-B6 and R2b-E8 libraries (HA2 20, 21, 41, 42, 44, 45 and 49) were all also flagged as mutational hotspots in the prior study.

There is also substantial agreement in the mutationally enriched positions in all 4 stem epitope antibody sorted libraries. This was an expected result given that these antibodies are clonally related. Mutations at the same set of residues, HA2 20, 21, 41, 42, 44, 45 and 49 are enriched for all 4 antibodies although the enrichment at position HA2 44 is much lower for the yeast display optimised antibodies. When mapped out on the structure of A/California/7/2009 HA (figure R4 5a,b,c,d), it appears that in the R2b-E8 sorted library and to a lesser extent the R1a-B6 sorted library, more residues peripheral to the core site are enriched compared to the WC11 and WC21 sorted libraries, indicating that WC11 and WC21 have potentially more compact epitopes.

6.2.4. Relating epitope mapping data to subtype sequence differences

To investigate the link between HA sequences of different subtypes and their relative ability to bind R1a-B6, R2b-E8, WC11 and WC21, the sequences of all the group 1 HA isolates used in experiments in this project were aligned and compared. Positions in the sequence which were flagged in Illumina sequencing data as enrichment hotspots for at least one of the stem binding antibodies are listed in table R4 3a along with enrichment factors for all antibodies and residues at the corresponding positions in all the HA isolates. It was reasoned that sequence polymorphisms responsible for differences in binding range would most likely be found at enrichment hotspots. The HA sequence at most of these positions was invariant but this search yielded 8 sequence polymorphisms compared to the sequence of A/California/7/2009 at 6 different positions (HA1 H38Q, V40K, V40Q, P299K, HA2 I45F, I45V, D46N, I48V). Enrichment factors for each of these individual mutations were calculated (table R4 3b). Due to the lower number of sorted cells in the R1a-B6 and WC11 sorted libraries and the low frequency in the unsorted library of the P299K mutation, calculated enrichment factors for these libraries and for this mutation were omitted.

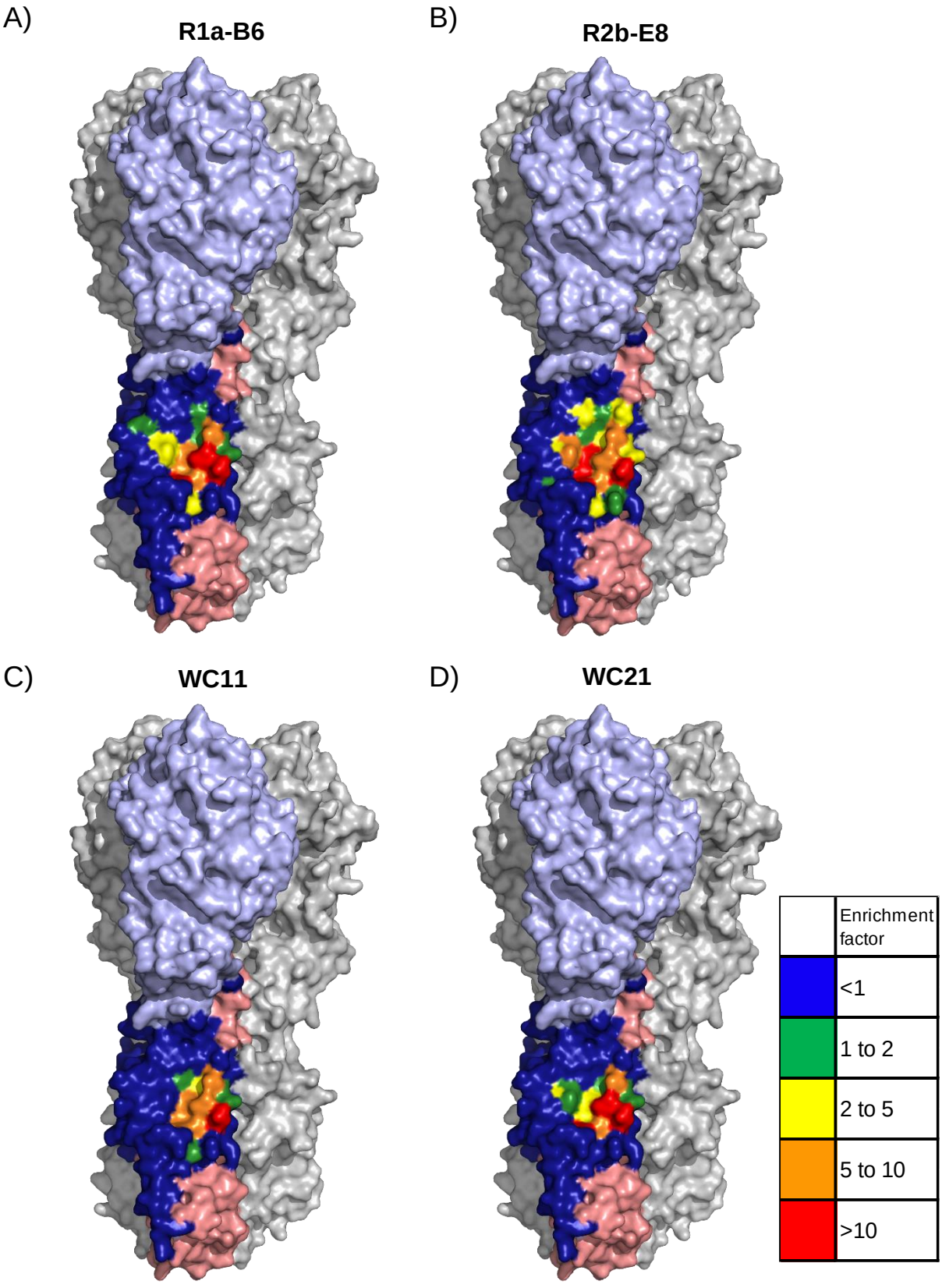


Figure R4 5a,b,c,d: Epitope heat maps of R1a-B6 clonal lineage VHH domains based on yeast display cell sorting Structures of H1 HA (A/California/7/2009) colour coded according to mutational enrichment factors according to the Illumina sequencing epitope mapping cell sorting data. Cell sorting enrichment factors were calculated by comparison of sequence data from the unsorted library to the libraries sorted for binding defects to **A)** R1a-B6 **B)** R2b-E8 **C)** WC11 **D)** WC21. Therefore, the greater the enrichment factor (EF), the more essential a residue is to the binding interaction. In the region sequenced in HA amplicons A and B, residues are coloured dark

blue (EF<1), green (EF1 to 2), yellow (EF2 to 5) orange (EF5-10) or red (EF>10) according to their enrichment factor. Residues not sequenced are coloured pale blue (HA1) or pale red (HA2). For all antibodies, a core set of residues in HA2 are all enriched 5-fold or greater (G20, T41, Q42, I45, T49). Several other residues are also consistently enriched. However, a larger number of residues peripheral to this core are enriched in the R1a-B6 library and a greater number still in the R2b-E8 sorted library.

A)

Residue number	Enrichment factor				Amino acids in different subtypes					
	R1a-B6	R2b-E8	WC11	WC21	H1	H2 A	H2 B	H5	H9 A	H9 B
HA1 26	0.0	1.9	4.3	2.3	V	V	V	V	V	V
HA1 28	0.0	0.0	6.9	0.2	T	T	T	T	T	T
HA1 37	3.4	3.4	0.0	2.1	T	T	T	T	T	T
HA1 38	2.9	7.5	0.1	1.3	H	H	H	H	H	Q
HA1 40	0.0	2.1	0.8	0.0	V	K	K	Q	K	K
HA1 299	2.1	0.4	0.0	0.3	P	P	P	P	K	K
HA1 317	0.0	1.5	2.9	1.2	A	A	A	A	A	A
HA1 319	3.6	2.6	0.6	3.7	G	G	G	G	G	G
HA2 20	12.6	12.2	9.2	14.6	G	G	G	G	G	G
HA2 21	7.2	12.0	8.9	4.0	W	W	W	W	W	W
HA2 22	2.1	4.3	1.8	0.9	Y	Y	Y	Y	Y	Y
HA2 23	5.5	8.9	6.0	3.7	G	G	G	G	G	G
HA2 36	2.2	2.7	1.1	0.1	A	A	A	A	A	A
HA2 41	6.3	8.4	9.4	8.0	T	T	T	T	T	T
HA2 42	11.5	12.3	18.8	13.9	Q	Q	Q	Q	Q	Q
HA2 44	10.9	10.6	2.6	3.1	A	A	A	A	A	A
HA2 45	13.2	9.3	9.7	15.2	I	F	F	I	I	V
HA2 46	1.2	3.1	1.4	1.2	D	D	N	D	D	D
HA2 48	1.1	4.6	4.6	1.9	I	I	I	V	I	I
HA2 49	5.6	6.5	6.3	6.3	T	T	T	T	T	T
HA2 53	0.0	4.4	0.0	0.0	N	N	N	N	N	N

B)

Mutation	Found in	Enrichment factor	
		R2b-E8	WC21
HA1 H38Q	H9	0.4	0.2
HA1 V40K	H2,H9	6.5	0.0
HA1 V40Q	H5	0.0	0.0
HA2 I45F	H2	12.4	0.4
HA2 I45V	H9	0.2	0.2
HA2 D46N	H2	0.8	0.1
HA2 I48V	H5	0.2	0.1

Table R4 3a,b: Using cell sorting enrichment factors to identify amino acid differences potentially responsible for subtype binding differences The left half of **A)** displays enrichment factor data for the residues assessed to be most involved in the binding interaction with one or more of the antibody panel (colour coding same as table R4 2a,b,c,d). The criterion for inclusion was an enrichment factor of 2 or greater in at least one of the sorted libraries. On the right half of the table are the amino acids at the equivalent positions in all the group 1 HA antigens used in this project. Antigens with identical amino acids at all these positions are grouped into one column: H1 (A/California/7/2009 and A/England/195/2009), H2 A (A/Canada/720/1957, A/England/1/1966, A/Singapore/1/1957 and A/Guiyang/1/1957), H2 B (A/Korea/426/1968), H5 (A/Anhui/1/2005, A/Vietnam/1194/2004 and A/Vietnam/1203/2004), H9 A (A/Hong Kong 1073/1999), H9 B (A/Chicken/Israel/2191417/2017). Amino acid differences when compared to A/California/7/2009 are highlighted in blue. **B)** The enrichment factors for these mutations were calculated to identify subtype sequence differences which may be responsible for binding differences to R1a-B6 related antibodies. HA1 V40K (present in H2 and H9 HA) and HA2 I45F (present in H2 HA) are heavily enriched in the R2b-E8 sorted library but are depleted in the WC21

sorted library. HA2 I45F has previously shown to abrogate R2b-E8 and R1a-B6 binding to H1 HA[102]. R2b-E8 has also shown poor H9 antigen binding and no H2 antigen binding in ELISA[103]. Both of these sequence differences may contribute to R2b-E8's inability to bind H2 antigen and HA1 V40K may explain poor binding to H9 HA. Conversely, if WC21 binding is unaffected by either mutation, this may partially explain H9 and H2 binding improvements. Due to the low frequency of each individual mutation, individual mutation enrichment factors from the R1a-B6 and WC11 sorted libraries and the HA1 P299K mutation were omitted. By multiplying the estimated frequency of a mutation in the unsorted library by the number of cells sorted, the number of sorted cells with a given mutation was estimated. If that number was under 10, the result was discarded.

Of these polymorphisms found in other subtypes, only HA1 V40K and HA2 I45F were enriched in the R2b-E8 sorted library and all mutations were depleted in the WC21 sorted library. HA1 V40K is found in all the H2 and H9 isolates used in this project and HA2 I45F is found in all the H2 isolates used. Previous data has shown that R2b-E8 binds poorly in ELISA to H9 antigen and binding to H2 antigen is undetectable[103]. Conversely, WC21 has shown strong binding to both (figure R3 8b,f). From these data, it was hypothesised that both HA1 V40K and HA2 I45F contribute to the observed differences in H2 binding to R2b-E8 and WC21 and HA1 V40K contributes to differences in H9 binding.

Several of the polymorphisms flagged in this deep mutational scanning experiment were previously tested on R1a-B6 and R2b-E8. These polymorphisms were HA2 I45F, I45V, D46N and I48V. Supporting the enrichment observed in this study, I45F HA did not bind to R2b-E8 (or R1a-B6). Additionally, D46N HA binding to R2b-E8 was not detected when tested individually in [103]. This mutation, however, was not enriched in the R2b-E8 sorted library as would be expected if there was no detectable binding.

6.2.5. Testing of individual HA mutants using flow cytometry

To experimentally test the hypotheses generated from the epitope mapping enrichment data, all 4 stem binding VHH were tested against yeast expressing HA with all 8 sequence polymorphisms. Yeast encoding HA with the mutations HA2 I45F, I45V, D46N and I48V were already available from the previous study and HA1 mutations H38Q, V40K, V40Q and P299K were inserted into pNIBS-5 encoding the A/California/7/2009 HA gene using the Quikchange II site directed mutagenesis kit (protocol and primers described in section 2.2.1).

Surface display of all 8 mutant HA (as well as WT HA) was induced in yeast and the yeast were stained with the 4 stem binding VHH (as well as R1a-G6 as a folding control) by the same protocol used prior to sorting (WC21 batch 2 was used which possessed a Myc-tag so yeast were stained with WC21 according to protocol 3a). For each VHH/HA combination, 2 x 10^5 cells were stained, and the staining protocol was scaled down appropriately. Fluorescence of the stained yeast populations was measured by flow cytometry and results are shown in table R4 4 and figure R4 6. As with the previous experiment, AF488 staining and AF647 staining correspond to HA display and VHH binding to HA respectively.

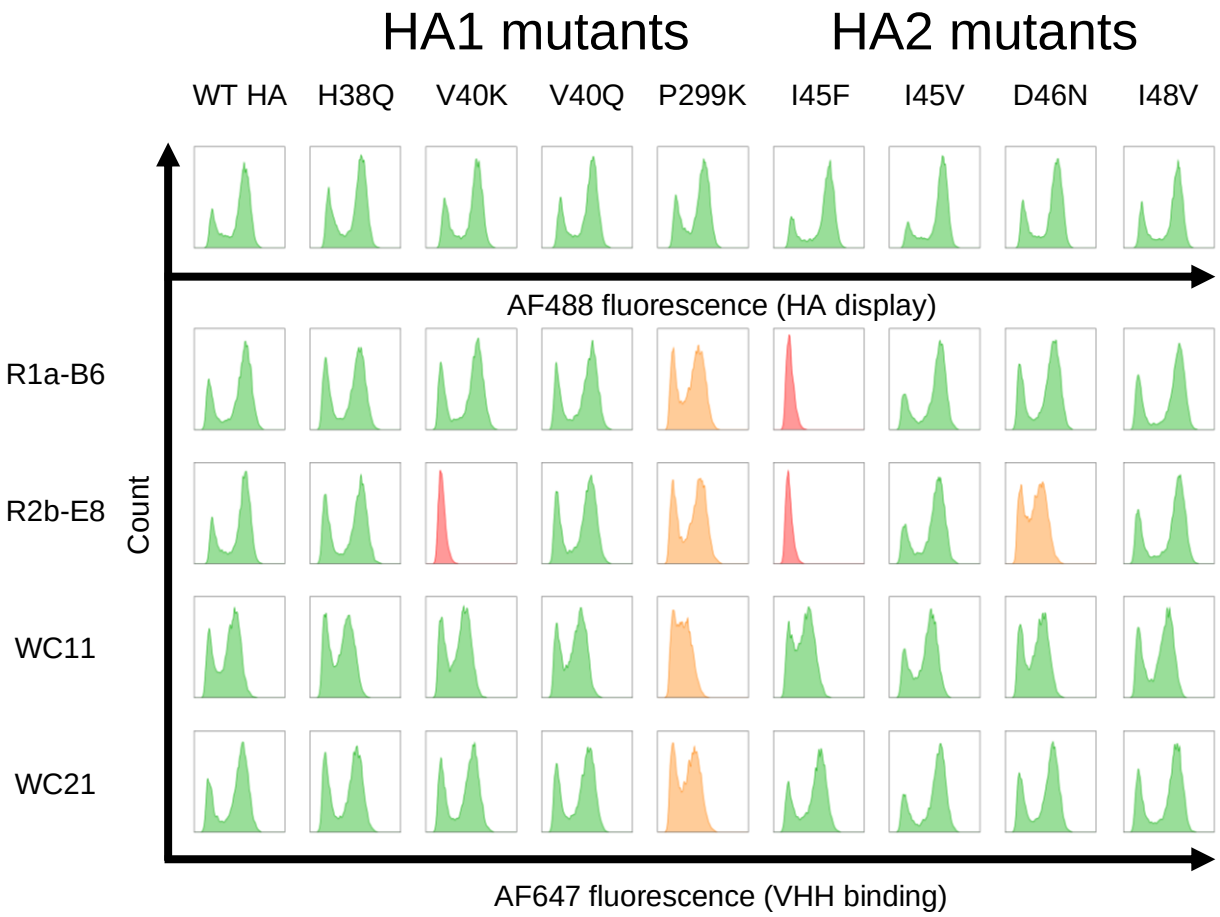


Figure R4 6: The effect of HA group 1 sequence polymorphisms on R1a-B6 clonal lineage antibody binding to HA Histograms are used to illustrate the HA display on and VHH binding to populations of HA mutant yeast as measured by flow cytometry. Mutants were picked based on FACS epitope mapping data and sequence comparison of group 1 HA antigens used in this project (see table R4 3a,b). AF488 fluorescence histograms (top row) show that all mutant variants of HA were expressed on yeast although HA1 P299K mutant was under-expressed compared to the others. Below are histograms of AF647 fluorescence, signifying VHH bound to expressed HA. VHH binding to HA in large quantity is signified by two well separated peaks and complete loss of binding of VHH can be observed as a single narrow peak. WC11 VHH can be seen to stain yeast less intensely, probably due to the partial loss of Myc-tag in this batch of antibody. Colouring of all histograms is based on AF488/AF647 MFI normalised as a percentage of the WT HA expressing yeast MFI (>50% - green, 50-10% - orange, <10% - red).

Mutation	Found in	AF647 MFI normailised to WT HA (%)			
		R1a-B6	R2b-E8	WC11	WC21
HA1 H38Q	H9	76	82	77	78
HA1 V40K	H2,H9	95	3	88	84
HA1 V40Q	H5	83	68	85	83
HA1 P299K	H9	34	37	34	32
HA2 I45F	H2	4	2	62	87
HA2 I45V	H9	95	81	122	132
HA2 D46N	H2	74	20	75	89
HA2 I48V	H5	147	143	138	153

Table R4 4: Comparison of HA yeast mutants binding to R1a-B6 clonal lineage antibodies
MFI measurements are used to compare the binding of mutants of H1 HA A/California/7/2009 displayed on yeast to VHH. All MFI values are displayed as a percentage of the MFI measurement of the corresponding VHH binding to WT H1 HA. Cells are colour coded according to MFI value (>50% - green, 50-10% - yellow, <10% - red). These values are used to highlight which VHH binding interactions are abrogated by sequence polymorphisms found in group 1 HA.

In line with predictions from the epitope mapping data, mutations HA1 V40K and HA2 I45F reduce R2b-E8 binding to undetectable levels and have little impact on WC21 binding. Similarly, WC11 binding is minimally impacted by either mutation. R1a-B6 binding is unaffected by HA1 V40K but as previously observed, binding is abrogated by HA2 I45F. In contrast to previous flow cytometry data, mutation HA2 D46N did show appreciable binding to R2b-E8, albeit significantly reduced compared to binding to WT HA. Weak but detectable binding to R2b-E8 would also explain why HA with HA2 D46N was neither strongly depleted nor enriched when sorted with R2b-E8. This mutation had little impact on HA binding to any of the other antibodies.

HA1 P299K appears to reduce AF647 MFI for all VHH approximately equally (~65%). In tandem, the AF488 MFI of this mutant yeast is reduced by approximately 40% compared to yeast expressing WT HA. All the other mutant yeast tested showed similar or greater AF488 staining compared to WT HA. It is therefore likely, that the low AF647 staining on HA HA1 P299K yeast is at least partially explained by reduced HA display.

Overall, the results of this epitope scan and testing of individual mutations present a partial model explaining many of the HA group 1 binding differences among the R1a-B6 clonal lineage. All of the stem epitope VHH have shown strong binding to H5 HA. Accordingly, the H5 sequence polymorphisms relative

to H1 identified in table R4 3a, HA1 V40Q and HA2 I48V, have little or no impact on binding.

ELISA data has shown that the binding of H9 HA to R2b-E8 is very weak compared to the other 3 VHH. The data indicates that a primary reason for this difference may be the HA1 40K residue found in H9 antigen which abrogates R2b-E8 binding but does not affect the other binding interactions.

Binding of H2 HA to R1a-B6 is weak and to R2b-E8 is undetectable. By contrast, WC11 and WC21 bind readily to H2 antigen. Correlating with this trend, the HA2 I45F mutations renders R1a-B6 and R2b-E8 binding to H1 HA undetectable but this mutation does not have a strong effect on WC11 and WC21 binding. It is likely that tolerance of this polymorphism is a primary reason that WC11 and WC21 can effectively bind H2 antigen.

The H2 antigens tested also possess the HA1 40K polymorphism, possibly explaining the fact that R2b-E8 binding to H2 is entirely undetectable by ELISA.

This data, however, does not suggest an explanation for the subtler difference in affinity for H9 HA observed between R1a-B6 and the affinity matured variants WC11 and WC21. Of the sequence polymorphisms tested, only HA1 V40K and P299K are found in the HA of H9-A/Hong Kong/1073/1999 and neither mutation was observed to have a differential effect on binding. It may be that the explanation for the observed affinity difference depends on more than one polymorphism in combination so cannot be detected with single mutants. Alternatively, casting a wider net by searching more of the H1-H9 sequence differences may have yielded polymorphisms which selectively diminish R1a-B6 binding compared to WC11 and WC21.

7. Conclusion and perspectives

This project aimed to identify and characterise variants of the broadly neutralising anti-influenza antibody R1a-B6 with a broader neutralisation profile than the unmodified antibody using *in vitro* antibody engineering methods. Initial characterisation of the R1a-B6/HA interaction was carried out with an alanine scan as well as identification and alignment of clonally related antibodies from Illumina sequencing data. Data from these experiments was used to design a set of 47 targeted amino acid substitutions at 20 different residue positions to introduce into R1a-B6 in a combinatorial library. This library was assembled using synthetic DNA, further mutagenised with error prone PCR and transformed into yeast for surface display. Illumina sequencing verified that the majority of the library was mutagenised as intended and assembled correctly.

Three rounds of sorting of the R1a-B6 variant library for improved binding to H9 HA yielded library H9R3b which, when compared to R1a-B6, stained 3-fold more strongly with H9 antigen and maintained strong staining with H1 and H5 antigen. Several R1a-B6 variants from H9 sorted libraries were selected for further characterisation. During the sorting process, R1a-B6 variants were isolated which bound directly to the streptavidin AF647 fluorophore used to label HA antigen as well as variants with mutations in their SV5-tag sequences used for normalisation. These results emphasised the importance of careful design of library sorting experiments to avoid unintentional enrichment of clones with unwanted properties, particularly in the context of sorting libraries against antigens with low initial binding activity.

Screening of the R1a-B6 variant clones isolated from H9 sorted libraries highlighted WC21. WC21 displayed a potent and even neutralisation profile against pseudoviruses with HA antigens from the pandemic risk subtypes H1, H2, H5 and H9. Compared to R1a-B6, WC21 was approximately 2 orders of magnitude more potent at neutralising H2 and H9 PV without any loss of potency against H1 and H5 PV. Investigation of WC21 using surface plasmon resonance and ELISA provided evidence that 3 amino acid substitutions in R1a-B6 were

primarily responsible for the affinity improvements observed in the WC21 variant. These 3 substitutions were R31I, D95N and Y102S.

Deep mutational scanning epitope mapping was used to characterise the footprints of R1a-B6 and its variants on the HA stem region. Data gathered from this experiment implied that the engineered R1a-B6 variants WC11 and WC21, compared to the parent antibody, rely on a smaller grouping of residues on HA for binding. We speculate that this may correlate with a broader range of antigen reactivity and more difficult viral escape. This experiment highlighted HA2 I45F and HA1 V40K as key determinants of antigenic differences between the stem regions of the subtypes of influenza A group 1. HA2 I45F, a polymorphism present on H2 antigens, blocked the binding of R1a-B6 to HA but WC11 and WC21 were able to overcome this barrier to binding.

Experiments examining the binding and neutralisation action of R1a-B6 and WC21 are ongoing. In a collaboration we have set up with Tracey Barrett (Department of Biological Sciences, Institute of Structural Molecular Biology, Birkbeck College, London, UK) we aim to generate 3-dimensional structures of the complexes formed by R1a-B6 and its variants with various HA molecules. One approach being tested is to use software backed up by the data collected in this thesis to model the interaction. The experimentally derived structures of the HA molecules used in these experiments are publicly available but structural information is not available for R1a-B6 or any of its variants. RoseTTAfold[226] and AlphaFold2[227] were used to model the structure of R1a-B6 from sequence and the resultant models were highly similar to each other. Several docking programs are being tested to combine HA and VHH structures into a complex. Some, such as HDOCK[228] assume that the protein models are rigid bodies while others such as HADDOCK[229] and AutoDock Vina[230] allow the models to flex in complex formation. Modelling pipelines will be tested by comparing models from sequence information with experimentally derived VHH/HA stem structures from the literature[108]. A major drawback associated with this approach is the fact that the CDR regions of VHH and of conventional antibodies are notoriously difficult to model owing to their high degree of flexibility. In addition, antibody-antigen complexes often involve a degree of induced fit. As an alternative, structural

studies of the R1aB6/HA complex are currently being pursued using electron microscopy.

Initial characterisation of the R1aB6/HA complex was carried out using negative stain electron microscopy. Samples were treated with heavy metal salts (uranyl acetate) prior to being imaged using a transmission electron microscope. Homogeneous species with threefold rotational symmetry were clearly observed. Samples were then flash frozen for more detailed analysis using cryogenic electron microscopy (Cryo-EM). Following the collection of 16,000 images, a subset (2000 images) was used for initial particle picking and the generation of 2D class averages from those similar in orientation. A selection of the 2D class averages generated are shown in figure CONC 1a,b,c,d. Encouragingly, all components of the HA are visible with additional density protruding from the stem region indicative of R1a-B6. Furthermore, secondary structure elements (alpha helices) can also be identified within the HA which is a positive indication that 3D reconstruction will be successful. This will require combination of the 2D class averages that represent 2D projections of the 3D structure and is currently underway[231, 232].

Structural data on the R1a-B6/HA interaction would provide atomic level explanations for the affinity differences caused by amino acid substitutions at the site of interaction both on the antibody side (R31I, D95N and Y102S) and on the HA side (HA2 I45F). By mutating residues within a structural model of a complex *in silico*, the effects of that mutation can be observed (e.g. Gain or loss of hydrogen bonds or introduction of a steric clash). The mechanisms by which R1a-B6 gains neutralisation breadth from amino acid substitutions will be compared to the mechanisms observed in other similar studies[138].

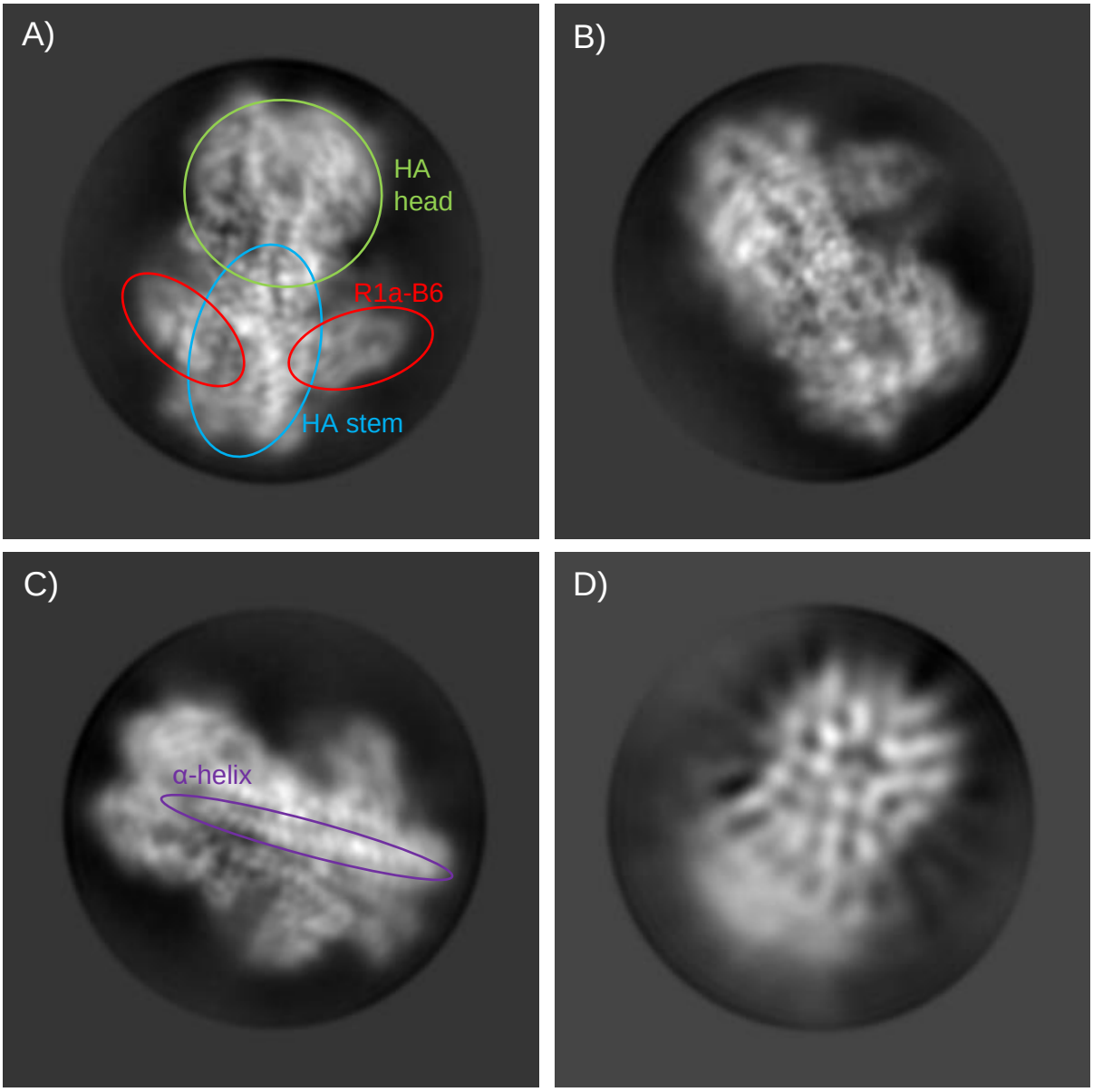


Figure CONC 1a,b,c,d: Preliminary negative stain Cryo-EM class average images of H1 HA in complex with R1a-B6 The complex consists of an H1 HA trimer (A/Brisbane/2/2018) and 3 R1a-B6 domains bound to each of the stem epitopes, giving the assembly a threefold rotational symmetry. **A)**, **B)** and **C)** appear to show the complex with the axis of symmetry perpendicular to the angle of view. The HA can be seen as a bulbous head region (highlighted in green) and a stem region (blue) consisting of a bundle of α-helices (purple). The protrusions from the stem are R1a-B6 molecules (red). **C)** is more difficult to interpret but may be an image of the complex from a top-down angle.

Attempts are also underway by Jason Long and Lethia Charles (Vaccines, Research and Development, Scientific Research and Innovation, Medicines and Healthcare Products Regulatory Agency, South Mimms, Potters Bar, UK) to run influenza live virus neutralisation assays comparing R1a-B6 and WC21. PV neutralisation titre is a well-established surrogate for live virus neutralisation. Although PV IC50 values tend to be lower, potent PV neutralisation is a strong

predictor of potent live virus neutralisation[215, 233]. Nonetheless, it is important to verify that the affinity and PV neutralisation potency improvements observed between R1a-B6 and WC21 translate into live virus neutralisation potency increases.

The data in this thesis suggests several ways that the project could be expanded. For example, all 3 of the driver mutations found in WC21 (R31I, D95N and Y102S) were found at high frequency in library H9R3b (20% or greater), supporting the notion that enriched mutations in this library are beneficial for HA binding. Several other individual mutations were enriched to a similar degree (R31S, S50R, P62S and D95A). The effects of each of these substitutions on affinity and neutralisation potency could be tested by applying them to R1a-B6 and to WC21 by site-directed mutagenesis. If available, an R1a-B6/HA complex structural model could be used to suggest additional substitutions to optimise the interaction (e.g. a substitution predicted to introduce a new hydrogen bond between antigen and antibody). In this way, further improvements could be made to the affinity and neutralisation range of R1a-B6. In particular, PV neutralisation assays indicated that R1a-B6 and WC21 neutralise H7 PV at very high concentrations. Further optimisation of WC21 may enhance its interaction with group 2 HA antigens.

The hypothesis that WC21 is more resistant to development of escape mutants than R1a-B6 could be tested with live virus. Epitope mapping data has provided profiles of HA amino acid substitutions capable of abrogating binding to R1a-B6 and WC21 but provides no information as to whether these HA mutations interfere with its biological function. Isolation of escape mutants highlights sequence changes compatible with the virus life cycle. Furthermore, escape mutations sometimes do not directly affect antibody binding or may be found on another viral protein gene[234-236]. Mutations such as these would not have been picked up in the experiments in this project.

To select for antibody escape mutants to R1a-B6 or WC21, virus could be passaged in cell culture in the presence of increasing concentrations of the antibody. Viruses could be sequenced to identify mutations showing signs of positive selection. The number of passages required for resistance mutations to

arise gives an indication of how likely resistance is to arise. One of the ways the reproductive fitness of resultant resistant viruses can be compared is using plaque assays. Fitter viruses tend to produce larger plaques on a cell monolayer so a smaller plaque can imply that resistance comes at a heavier cost to fitness[236]. More passages or less fit escape mutants would imply that WC21 is more resistant to viral escape.

The emergence of the Omicron strain of SARS-CoV2 has highlighted the importance of developing monoclonal antibodies which are difficult to evolve escape mutants to. Most of the monoclonal antibodies approved for use against COVID19 have lost potency or are no longer neutralising against currently circulating viruses[237]. Monoclonal antibodies which resist the development of escape mutant viral stains will likely have longer lifespans as diagnostic reagents and therapeutics.

The competitive ELISA based vaccine potency assay described in [142] using R1a-B6 could potentially be improved by substituting R1a-B6 for WC21. We have demonstrated that WC21 binding is abrogated by narrower range of mutations than R1a-B6. Therefore, we believe that it is less likely that a future strains of H1 seasonal influenza would lose binding to WC21 compared to R1a-B6, potentially giving the assay a longer lifespan before new single domain antibodies have to be isolated as replacements. The paper also described application of the same potency assay to quantification of native H5 HA. The use of WC21 in this assay would allow quantification of H2 and H9 in addition to H1 and H5 HA antigen.

Further work could be done on defining in more detail the binding and neutralisation range of WC21. Only 6 subtypes of HA have been tested against the VHH (H1, H2, H3, H5, H7 and H9). Of each of these subtypes, HA from only 1 or a few isolates have been used in experiments. ELISA would be the best format to test a variety of influenza isolates for WC21 binding due to the availability of a wide range of lyophilised influenza antigen standards. By comparing isolate HA sequence with epitope mapping enrichment data and ELISA EC50 values, the extent to which epitope mapping data is predictive of binding could be ascertained. We hypothesise that single amino acid polymorphisms to H1 (A/California/7/2009) identified in epitope mapping

enrichment data as abrogating WC21 binding, would also abrogate WC21 binding against the antigenic background of different HA subtypes[102]. If this hypothesis is verified, retention of binding of WC21 to any newly emerging influenza isolates could be predicted from sequence alone.

In [141] R1a-B6 was used in an AAV-vectored immunotherapy for influenza prophylaxis. In this application, substitution of R1a-B6 for the improved variant WC21 would expand the utility of this concept. After AAV treatment, the production of antibody is intended to be permanent or semi-permanent so the antibody in question must be highly likely to neutralise future seasonal and pandemic strains of influenza. In addition to the H1 and H5 influenza strains R1a-B6 is protective against, a WC21-based AAV immunotherapy would also likely be protective against H2 and H9 influenza viruses if these subtypes were to circulate in the human population in the future. In this application, due to long-term presence in the human bloodstream, WC21 would have to be humanised to resemble a human VH domain to minimise the chances of immunogenicity or any other adverse responses[137].

An alternative option for prophylactic use of WC21 is intranasal delivery. Several proof-of-concept studies and drugs in development use the modularity and ease of expression of VHH for intranasal protection against respiratory pathogens including influenza[127, 238, 239]. These constructs can be expressed in and purified from yeast rather than mammalian cells, reducing cost, an important consideration for a short-term prophylactic. Furthermore, humanisation may not be necessary for this delivery at this site[240], avoiding the loss of affinity often observed during this process[137].

The output of this project is a VHH antibody with a substantially broadened binding and neutralisation range compared to the parental antibody. The affinity improvements achieved make WC21 more attractive for a range of seasonal influenza and pandemic preparedness applications. The identification of the amino acid changes in WC21 responsible for functional improvements and the inter-subtype polymorphisms responsible for stem epitope antigenic differences begins the process of a molecular level understanding of this antibody-antigen interaction.

Furthermore, the methods used in this thesis provide an additional case study for the use of antibody engineering techniques to broaden the range of antigen variants an antibody can bind. Comparatively few studies have been published on this application of antibody engineering. There is a continual need for broadly active, prophylactic and therapeutic drugs against a range of pathogens with antigenic variability such as HIV, hepatis C virus and SARS-CoV2[241, 242]. *In vitro* cross-reactivity optimised VHH domains could substantially contribute to reducing the burden of these diseases. In addition, similar techniques could be applied to related fields such as the production of hapten binding antibodies which target whole families of small molecules[243, 244].

8. References

1. Saunders-Hastings, P.R. and D. Krewski, *Reviewing the History of Pandemic Influenza: Understanding Patterns of Emergence and Transmission.* Pathogens, 2016. **5**(4): p. 66.
2. Mamelund, S.E., *Influenza, historical.* Medicine, 2008. **54**: p. 361-371.
3. Paget, J., et al., *Global mortality associated with seasonal influenza epidemics: New burden estimates and predictors from the GLaMOR Project.* J Glob Health, 2019. **9**(2): p. 020421.
4. Iuliano, A.D., et al., *Estimates of global seasonal influenza-associated respiratory mortality: a modelling study.* Lancet, 2018. **391**(10127): p. 1285-1300.
5. Smith, D.J., et al., *Mapping the antigenic and genetic evolution of influenza virus.* Science, 2004. **305**(5682): p. 371-6.
6. Viboud, C., et al., *Global Mortality Impact of the 1957-1959 Influenza Pandemic.* J Infect Dis, 2016. **213**(5): p. 738-45.
7. Gupta, V., et al., *Socioeconomic impact due to COVID-19: An empirical assessment.* Inf Process Manag, 2022. **59**(2): p. 102810.
8. Lefkowitz, E.J., et al., *Virus taxonomy: the database of the International Committee on Taxonomy of Viruses (ICTV).* Nucleic Acids Res, 2018. **46**(D1): p. D708-D717.
9. Walker, P.J., et al., *Recent changes to virus taxonomy ratified by the International Committee on Taxonomy of Viruses (2022).* Arch Virol, 2022. **167**(11): p. 2429-2440.
10. Su, S., et al., *Novel Influenza D virus: Epidemiology, pathology, evolution and biological characteristics.* Virulence, 2017. **8**(8): p. 1580-1591.
11. Furuse, Y., et al., *Analyses of Evolutionary Characteristics of the Hemagglutinin-Esterase Gene of Influenza C Virus during a Period of 68 Years Reveals Evolutionary Patterns Different from Influenza A and B Viruses.* Viruses, 2016. **8**(12): p. 321.
12. Zaraket, H., et al., *Burden of influenza B virus infection and considerations for clinical management.* Antiviral Res, 2021. **185**: p. 104970.
13. Kaplan, B.S. and R.J. Webby, *The avian and mammalian host range of highly pathogenic avian H5N1 influenza.* Virus Res, 2013. **178**(1): p. 3-11.
14. Kessler, S., et al., *Influenza A Viruses and Zoonotic Events-Are We Creating Our Own Reservoirs?* Viruses, 2021. **13**(11): p. 2250.
15. Skehel, J.J. and D.C. Wiley, *Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin.* Annu Rev Biochem, 2000. **69**: p. 531-69.
16. Dou, D., et al., *Influenza A Virus Cell Entry, Replication, Virion Assembly and Movement.* Front Immunol, 2018. **9**: p. 1581.
17. Rust, M.J., et al., *Assembly of endocytic machinery around individual influenza viruses during viral entry.* Nat Struct Mol Biol, 2004. **11**(6): p. 567-73.
18. Bui, M., G. Whittaker, and A. Helenius, *Effect of M1 protein and low pH on nuclear transport of influenza virus ribonucleoproteins.* J Virol, 1996. **70**(12): p. 8391-401.
19. Boonstra, S., et al., *Hemagglutinin-Mediated Membrane Fusion: A Biophysical Perspective.* Annu Rev Biophys, 2018. **47**: p. 153-173.
20. De Vlugt, C., D. Sikora, and M. Pelchat, *Insight into Influenza: A Virus Cap-Snatching.* Viruses, 2018. **10**(11): p. 641.
21. Fodor, E. and A.J.W. Te Velthuis, *Structure and Function of the Influenza Virus Transcription and Replication Machinery.* Cold Spring Harb Perspect Med, 2020. **10**(9): p. a038398.
22. Rossman, J.S. and R.A. Lamb, *Influenza virus assembly and budding.* Virology, 2011. **411**(2): p. 229-36.
23. Motsa, B.B. and R.V. Stahelin, *Lipid-protein interactions in virus assembly and budding from the host cell plasma membrane.* Biochem Soc Trans, 2021. **49**(4): p. 1633-1641.

24. Benton, D.J., et al., *Structural transitions in influenza haemagglutinin at membrane fusion pH.* Nature, 2020. **583**(7814): p. 150-153.
25. Garcia-Moro, E., et al., *Reversible structural changes in the influenza hemagglutinin precursor at membrane fusion pH.* Proc Natl Acad Sci U S A, 2022. **119**(33): p. e2208011119.
26. Wu, N.C. and I.A. Wilson, *A Perspective on the Structural and Functional Constraints for Immune Evasion: Insights from Influenza Virus.* J Mol Biol, 2017. **429**(17): p. 2694-2709.
27. Smith, G.J., et al., *Dating the emergence of pandemic influenza viruses.* Proc Natl Acad Sci U S A, 2009. **106**(28): p. 11709-12.
28. Tong, S., et al., *New world bats harbor diverse influenza A viruses.* PLoS Pathog, 2013. **9**(10): p. e1003657.
29. Xu, R., et al., *Structural basis of preexisting immunity to the 2009 H1N1 pandemic influenza virus.* Science, 2010. **328**(5976): p. 357-60.
30. Zhang, Y., et al., *Influenza Research Database: An integrated bioinformatics resource for influenza virus research.* Nucleic Acids Res, 2017. **45**(D1): p. D466-D474.
31. Tung Yep, A., et al., *Broad Reactivity Single Domain Antibodies against Influenza Virus and Their Applications to Vaccine Potency Testing and Immunotherapy.* Biomolecules, 2021. **11**(3): p. 407.
32. Noh, J.Y. and W.J. Kim, *Influenza vaccines: unmet needs and recent developments.* Infect Chemother, 2013. **45**(4): p. 375-86.
33. Harrington, W.N., C.M. Kackos, and R.J. Webby, *The evolution and future of influenza pandemic preparedness.* Exp Mol Med, 2021. **53**(5): p. 737-749.
34. Neumann, G. and Y. Kawaoka, *Predicting the Next Influenza Pandemics.* J Infect Dis, 2019. **219**(Suppl_1): p. S14-S20.
35. Smith, G.J., et al., *Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic.* Nature, 2009. **459**(7250): p. 1122-5.
36. Harding, A.T. and N.S. Heaton, *Efforts to Improve the Seasonal Influenza Vaccine.* Vaccines (Basel), 2018. **6**(2).
37. Sona, S., et al., *Influenza virus and its subtypes circulating during 2018-2019: A hospital-based study from Assam.* Indian J Med Microbiol, 2022: p. 525-530.
38. Tscherne, D.M. and A. Garcia-Sastre, *Virulence determinants of pandemic influenza viruses.* J Clin Invest, 2011. **121**(1): p. 6-13.
39. Cheung, J.T.L., et al., *Determining Existing Human Population Immunity as Part of Assessing Influenza Pandemic Risk.* Emerg Infect Dis, 2022. **28**(5): p. 977-985.
40. Babu, T.M., et al., *Population Serologic Immunity to Human and Avian H2N2 Viruses in the United States and Hong Kong for Pandemic Risk Assessment.* J Infect Dis, 2018. **218**(7): p. 1054-1060.
41. Kim, S.H., *Challenge for One Health: Co-Circulation of Zoonotic H5N1 and H9N2 Avian Influenza Viruses in Egypt.* Viruses, 2018. **10**(3): p. 121.
42. Oshansky, C.M., et al., *Adjuvanted recombinant hemagglutinin H7 vaccine to highly pathogenic influenza A(H7N9) elicits high and sustained antibody responses in healthy adults.* NPJ Vaccines, 2021. **6**(1): p. 41.
43. Khan, S.U., et al., *A Systematic Review and Meta-Analysis of the Seroprevalence of Influenza A(H9N2) Infection Among Humans.* J Infect Dis, 2015. **212**(4): p. 562-9.
44. Bi, Y., J. Li, and W. Shi, *The time is now: a call to contain H9N2 avian influenza viruses.* Lancet Microbe, 2022: p. e804-e805.
45. Sun, Y. and J. Liu, *H9N2 influenza virus in China: a cause of concern.* Protein Cell, 2015. **6**(1): p. 18-25.
46. Air, G.M., *Influenza virus antigenicity and broadly neutralizing epitopes.* Curr Opin Virol, 2015. **11**: p. 113-21.

47. Joyce, M.G., et al., *Vaccine-Induced Antibodies that Neutralize Group 1 and Group 2 Influenza A Viruses.* Cell, 2016. **166**(3): p. 609-623.
48. Shriver, Z., J.M. Trevejo, and R. Sasisekharan, *Antibody-Based Strategies to Prevent and Treat Influenza.* Front Immunol, 2015. **6**: p. 315.
49. Rolfes, M.A., et al., *Annual estimates of the burden of seasonal influenza in the United States: A tool for strengthening influenza surveillance and preparedness.* Influenza Other Respir Viruses, 2018. **12**(1): p. 132-137.
50. Paules, C.I. and A.S. Fauci, *Influenza Vaccines: Good, but We Can Do Better.* J Infect Dis, 2019. **219**(Suppl_1): p. S1-S4.
51. Gouma, S., E.M. Anderson, and S.E. Hensley, *Challenges of Making Effective Influenza Vaccines.* Annu Rev Virol, 2020. **7**(1): p. 495-512.
52. Coudeville, L., et al., *Relationship between haemagglutination-inhibiting antibody titres and clinical protection against influenza: development and application of a bayesian random-effects model.* BMC Med Res Methodol, 2010. **10**: p. 18.
53. Caton, A.J., et al., *The antigenic structure of the influenza virus A/PR/8/34 hemagglutinin (H1 subtype).* Cell, 1982. **31**(2 Pt 1): p. 417-27.
54. Gerhard, W., et al., *Antigenic structure of influenza virus haemagglutinin defined by hybridoma antibodies.* Nature, 1981. **290**(5808): p. 713-7.
55. Wiley, D.C., I.A. Wilson, and J.J. Skehel, *Structural identification of the antibody-binding sites of Hong Kong influenza haemagglutinin and their involvement in antigenic variation.* Nature, 1981. **289**(5796): p. 373-8.
56. Knossow, M., et al., *Mechanism of neutralization of influenza virus infectivity by antibodies.* Virology, 2002. **302**(2): p. 294-8.
57. Brandenburg, B., et al., *Mechanisms of hemagglutinin targeted influenza virus neutralization.* PLoS One, 2013. **8**(12): p. e80034.
58. Ohmit, S.E., et al., *Influenza hemagglutination-inhibition antibody titer as a correlate of vaccine-induced protection.* J Infect Dis, 2011. **204**(12): p. 1879-85.
59. Gamblin, S.J. and J.J. Skehel, *Influenza hemagglutinin and neuraminidase membrane glycoproteins.* J Biol Chem, 2010. **285**(37): p. 28403-9.
60. Tricco, A.C., et al., *Comparing influenza vaccine efficacy against mismatched and matched strains: a systematic review and meta-analysis.* BMC Med, 2013. **11**: p. 153.
61. Cowling, B.J., et al., *Protective efficacy of seasonal influenza vaccination against seasonal and pandemic influenza virus infection during 2009 in Hong Kong.* Clin Infect Dis, 2010. **51**(12): p. 1370-9.
62. Cowling, B.J., et al., *Protective efficacy against pandemic influenza of seasonal influenza vaccination in children in Hong Kong: a randomized controlled trial.* Clin Infect Dis, 2012. **55**(5): p. 695-702.
63. Goodwin, K., C. Viboud, and L. Simonsen, *Antibody response to influenza vaccination in the elderly: a quantitative review.* Vaccine, 2006. **24**(8): p. 1159-69.
64. Smetana, J., et al., *Influenza vaccination in the elderly.* Hum Vaccin Immunother, 2018. **14**(3): p. 540-549.
65. Okuno, Y., et al., *A common neutralizing epitope conserved between the hemagglutinins of influenza A virus H1 and H2 strains.* J Virol, 1993. **67**(5): p. 2552-8.
66. Dreyfus, C., D.C. Ekiert, and I.A. Wilson, *Structure of a classical broadly neutralizing stem antibody in complex with a pandemic H2 influenza virus hemagglutinin.* J Virol, 2013. **87**(12): p. 7149-54.
67. Ekiert, D.C., et al., *Antibody recognition of a highly conserved influenza virus epitope.* Science, 2009. **324**(5924): p. 246-51.
68. Throsby, M., et al., *Heterosubtypic neutralizing monoclonal antibodies cross-protective against H5N1 and H1N1 recovered from human IgM+ memory B cells.* PLoS One, 2008. **3**(12): p. e3942.

69. Sui, J., et al., *Structural and functional bases for broad-spectrum neutralization of avian and human influenza A viruses.* Nat Struct Mol Biol, 2009. **16**(3): p. 265-73.
70. Corti, D., et al., *A neutralizing antibody selected from plasma cells that binds to group 1 and group 2 influenza A hemagglutinins.* Science, 2011. **333**(6044): p. 850-6.
71. Kallewaard, N.L., et al., *Structure and Function Analysis of an Antibody Recognizing All Influenza A Subtypes.* Cell, 2016. **166**(3): p. 596-608.
72. Lang, S., et al., *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 Rep, 2017. **20**(12): p. 2935-2943.
73. Dreyfus, C., et al., *Highly conserved protective epitopes on influenza B viruses.* Science, 2012. **337**(6100): p. 1343-8.
74. Friesen, R.H., et al., *A common solution to group 2 influenza virus neutralization.* Proc Natl Acad Sci U S A, 2014. **111**(1): p. 445-50.
75. Nakamura, G., et al., *An in vivo human-plasmablast enrichment technique allows rapid identification of therapeutic influenza A antibodies.* Cell Host Microbe, 2013. **14**(1): p. 93-103.
76. McCarthy, K.R., et al., *Memory B Cells that Cross-React with Group 1 and Group 2 Influenza A Viruses Are Abundant in Adult Human Repertoires.* Immunity, 2018. **48**(1): p. 174-184 e9.
77. Ekiert, D.C., et al., *Cross-neutralization of influenza A viruses mediated by a single antibody loop.* Nature, 2012. **489**(7417): p. 526-32.
78. Zhu, X., et al., *Influenza chimeric hemagglutinin structures in complex with broadly protective antibodies to the stem and trimer interface.* Proc Natl Acad Sci U S A, 2022. **119**(21): p. e2200821119.
79. Bajic, G., et al., *Influenza Antigen Engineering Focuses Immune Responses to a Subdominant but Broadly Protective Viral Epitope.* Cell Host Microbe, 2019. **25**(6): p. 827-835 e6.
80. Watanabe, A., et al., *Antibodies to a Conserved Influenza Head Interface Epitope Protect by an IgG Subtype-Dependent Mechanism.* Cell, 2019. **177**(5): p. 1124-1135 e16.
81. Doyle, T.M., et al., *Universal anti-neuraminidase antibody inhibiting all influenza A subtypes.* Antiviral Res, 2013. **100**(2): p. 567-74.
82. Wang, R., et al., *Therapeutic potential of a fully human monoclonal antibody against influenza A virus M2 protein.* Antiviral Res, 2008. **80**(2): p. 168-77.
83. Wang, Y., et al., *Monoclonal antibody recognizing SLLTEVET epitope of M2 protein potently inhibited the replication of influenza A viruses in MDCK cells.* Biochem Biophys Res Commun, 2009. **385**(1): p. 118-22.
84. DiLillo, D.J., et al., *Broadly neutralizing hemagglutinin stalk-specific antibodies require FcgammaR interactions for protection against influenza virus in vivo.* Nat Med, 2014. **20**(2): p. 143-51.
85. Narayan, R. and S. Tripathi, *Intrinsic ADE: The Dark Side of Antibody Dependent Enhancement During Dengue Infection.* Front Cell Infect Microbiol, 2020. **10**: p. 580096.
86. Dejnirattisai, W., et al., *Cross-reacting antibodies enhance dengue virus infection in humans.* Science, 2010. **328**(5979): p. 745-8.
87. Chan-Hui, P.Y. and K.M. Swiderek, *Immunological considerations for developing antibody therapeutics for Influenza A.* Hum Vaccin Immunother, 2016. **12**(2): p. 474-7.
88. Ramakrishnan, B., et al., *A Structural and Mathematical Modeling Analysis of the Likelihood of Antibody-Dependent Enhancement in Influenza.* Trends Microbiol, 2016. **24**(12): p. 933-943.

89. Khurana, S., et al., *Vaccine-induced anti-HA2 antibodies promote virus fusion and enhance influenza virus respiratory disease.* Sci Transl Med, 2013. **5**(200): p. 200ra114.
90. Winarski, K.L., et al., *Antibody-dependent enhancement of influenza disease promoted by increase in hemagglutinin stem flexibility and virus fusion kinetics.* Proc Natl Acad Sci U S A, 2019. **116**(30): p. 15194-15199.
91. Fox, A., K.M. Quinn, and K. Subbarao, *Extending the Breadth of Influenza Vaccines: Status and Prospects for a Universal Vaccine.* Drugs, 2018. **78**(13): p. 1297-1308.
92. Sparrow, E., et al., *Passive immunization for influenza through antibody therapies, a review of the pipeline, challenges and potential applications.* Vaccine, 2016. **34**(45): p. 5442-5448.
93. Koszalka, P., K. Subbarao, and M. Baz, *Preclinical and clinical developments for combination treatment of influenza.* PLoS Pathog, 2022. **18**(5): p. e1010481.
94. Behzadi, M.A. and V.H. Leyva-Grado, *Overview of Current Therapeutics and Novel Candidates Against Influenza, Respiratory Syncytial Virus, and Middle East Respiratory Syndrome Coronavirus Infections.* Front Microbiol, 2019. **10**: p. 1327.
95. Corbett, K.S., et al., *Design of Nanoparticulate Group 2 Influenza Virus Hemagglutinin Stem Antigens That Activate Unmutated Ancestor B Cell Receptors of Broadly Neutralizing Antibody Lineages.* mBio, 2019. **10**(1): p. e02810-18.
96. Wang, W.C., et al., *Progress towards the Development of a Universal Influenza Vaccine.* Viruses, 2022. **14**(8): p. 1684.
97. Bullard, B.L. and E.A. Weaver, *Strategies Targeting Hemagglutinin as a Universal Influenza Vaccine.* Vaccines (Basel), 2021. **9**(3); p. 257.
98. Chu, J.T., et al., *Heterosubtypic immune pressure accelerates emergence of influenza A virus escape phenotypes in mice.* Virus Res, 2022. **323**: p. 198991.
99. Lee, P.S., et al., *Receptor mimicry by antibody F045-092 facilitates universal binding to the H3 subtype of influenza virus.* Nat Commun, 2014. **5**: p. 3614.
100. Zhou, T., et al., *Structural Repertoire of HIV-1-Neutralizing Antibodies Targeting the CD4 Supersite in 14 Donors.* Cell, 2015. **161**(6): p. 1280-92.
101. Kwong, P.D. and I.A. Wilson, *HIV-1 and influenza antibodies: seeing antigens in new ways.* Nat Immunol, 2009. **10**(6): p. 573-8.
102. Gaiotto, T. and S.E. Hufton, *Cross-Neutralising Nanobodies Bind to a Conserved Pocket in the Hemagglutinin Stem Region Identified Using Yeast Display and Deep Mutational Scanning.* PLoS One, 2016. **11**(10): p. e0164296.
103. Hufton, S.E., et al., *The breadth of cross sub-type neutralisation activity of a single domain antibody to influenza hemagglutinin can be increased by antibody valency.* PLoS One, 2014. **9**(8): p. e103294.
104. Lerner, R.A., *Rare antibodies from combinatorial libraries suggests an S.O.S. component of the human immunological repertoire.* Mol Biosyst, 2011. **7**(4): p. 1004-12.
105. Muyldermans, S., et al., *Sequence and structure of VH domain from naturally occurring camel heavy chain immunoglobulins lacking light chains.* Protein Eng, 1994. **7**(9): p. 1129-35.
106. Hamers-Casterman, C., et al., *Naturally occurring antibodies devoid of light chains.* Nature, 1993. **363**(6428): p. 446-8.
107. Muyldermans, S., *Nanobodies: natural single-domain antibodies.* Annu Rev Biochem, 2013. **82**: p. 775-97.
108. Laursen, N.S., et al., *Universal protection against influenza infection by a multidomain antibody to influenza hemagglutinin.* Science, 2018. **362**(6414): p. 598-602.

109. Kabat, E.A., Wu T. T., Bilofsky, H., *Sequences of immunoglobulin chains: tabulation and analysis of amino acid sequences of precursors, V-regions, C-regions, J-chain and 2-microglobulins.* National Institute of Health, 1979.
110. Tu, Z., et al., *Landscape of variable domain of heavy-chain-only antibody repertoire from alpaca.* Immunology, 2020. **161**(1): p. 53-65.
111. Govaert, J., et al., *Dual beneficial effect of interloop disulfide bond for single domain antibody fragments.* J Biol Chem, 2012. **287**(3): p. 1970-9.
112. Gonzalez-Sapienza, G., M.A. Rossotti, and S. Tabares-da Rosa, *Single-Domain Antibodies As Versatile Affinity Reagents for Analytical and Diagnostic Applications.* Front Immunol, 2017. **8**: p. 977.
113. Mitchell, L.S. and L.J. Colwell, *Comparative analysis of nanobody sequence and structure data.* Proteins, 2018. **86**(7): p. 697-706.
114. Mitchell, L.S. and L.J. Colwell, *Analysis of nanobody paratopes reveals greater diversity than classical antibodies.* Protein Eng Des Sel, 2018. **31**(7-8): p. 267-275.
115. Henry, K.A. and C.R. MacKenzie, *Antigen recognition by single-domain antibodies: structural latitudes and constraints.* MAbs, 2018. **10**(6): p. 815-826.
116. De Genst, E., et al., *Molecular basis for the preferential cleft recognition by dromedary heavy-chain antibodies.* Proc Natl Acad Sci U S A, 2006. **103**(12): p. 4586-91.
117. Hendrickx, M.L., et al., *TAFIa inhibiting nanobodies as profibrinolytic tools and discovery of a new TAFIa conformation.* J Thromb Haemost, 2011. **9**(11): p. 2268-77.
118. Henry, K.A. and C.R. MacKenzie, *Editorial: Single-Domain Antibodies-Biology, Engineering and Emerging Applications.* Front Immunol, 2018. **9**: p. 41.
119. Stijlemans, B., et al., *Efficient targeting of conserved cryptic epitopes of infectious agents by single domain antibodies. African trypanosomes as paradigm.* J Biol Chem, 2004. **279**(2): p. 1256-61.
120. Liu, Y. and H. Huang, *Expression of single-domain antibody in different systems.* Appl Microbiol Biotechnol, 2018. **102**(2): p. 539-551.
121. Shkoporov, A.N., et al., *Production of biologically active scFv and VHH antibody fragments in Bifidobacterium longum.* FEMS Microbiol Lett, 2015. **362**(12): p. fnv083.
122. Ramage, W., et al., *Cross-Reactive and Lineage-Specific Single Domain Antibodies against Influenza B Hemagglutinin.* Antibodies (Basel), 2019. **8**(1): p. 14.
123. Gaiotto, T., et al., *Nanobodies mapped to cross-reactive and divergent epitopes on A(H7N9) influenza hemagglutinin using yeast display.* Sci Rep, 2021. **11**(3126).
124. Tillib, S.V., et al., *Formatted single-domain antibodies can protect mice against infection with influenza virus (H5N2).* Antiviral Res, 2013. **97**(3): p. 245-54.
125. Ibanez, L.I., et al., *Nanobodies with in vitro neutralizing activity protect mice against H5N1 influenza virus infection.* J Infect Dis, 2011. **203**(8): p. 1063-72.
126. Cardoso, F.M., et al., *Single-domain antibodies targeting neuraminidase protect against an H5N1 influenza virus challenge.* J Virol, 2014. **88**(15): p. 8278-96.
127. De Vlieger, D., et al., *Selective Engagement of FcgammaRIV by a M2e-Specific Single Domain Antibody Construct Protects Against Influenza A Virus Infection.* Front Immunol, 2019. **10**: p. 2920.
128. Wei, G., et al., *Potent neutralization of influenza A virus by a single-domain antibody blocking M2 ion channel protein.* PLoS One, 2011. **6**(12): p. e28309.
129. Panza, P., et al., *Live imaging of endogenous protein dynamics in zebrafish using chromobodies.* Development, 2015. **142**(10): p. 1879-84.
130. Liu, J.L., et al., *Bioconjugates of rhizavidin with single domain antibodies as bifunctional immunoreagents.* J Immunol Methods, 2014. **411**: p. 37-42.

131. Koide, A., et al., *Exploring the capacity of minimalist protein interfaces: interface energetics and affinity maturation to picomolar KD of a single-domain antibody with a flat paratope.* J Mol Biol, 2007. **373**(4): p. 941-53.
132. De Genst, E., et al., *Chemical basis for the affinity maturation of a camel single domain antibody.* J Biol Chem, 2004. **279**(51): p. 53593-601.
133. Hu, M., et al., *In vitro affinity maturation to improve the efficacy of a hypoxia-inducible factor 1alpha single-domain intrabody.* Biochem Biophys Res Commun, 2020. **529**(4): p. 936-942.
134. Ikeuchi, E., et al., *Delicate balance among thermal stability, binding affinity, and conformational space explored by single-domain V(H)H antibodies.* Sci Rep, 2021. **11**(1): p. 20624.
135. Yogi, A., et al., *Brain Delivery of IGF1R5, a Single-Domain Antibody Targeting Insulin-like Growth Factor-1 Receptor.* Pharmaceutics, 2022. **14**(7): p. e22208.
136. Sulea, T., *Humanization of Camelid Single-Domain Antibodies.* Methods Mol Biol, 2022. **2446**: p. 299-312.
137. Rossotti, M.A., et al., *Immunogenicity and humanization of single-domain antibodies.* FEBS J, 2022. **289**(14): p. 4304-4327.
138. Sulea, T., et al., *Structure-based dual affinity optimization of a SARS-CoV-1/2 cross-reactive single-domain antibody.* PLoS One, 2022. **17**(3): p. e0266250.
139. Boder, E.T., K.S. Midelfort, and K.D. Wittrup, *Directed evolution of antibody fragments with monovalent femtomolar antigen-binding affinity.* Proc Natl Acad Sci U S A, 2000. **97**(20): p. 10701-5.
140. Lopez, T., et al., *Epitope-specific affinity maturation improved stability of potent protease inhibitory antibodies.* Biotechnol Bioeng, 2018. **115**(11): p. 2673-2682.
141. Del Rosario, J.M.M., et al., *Protection From Influenza by Intramuscular Gene Vector Delivery of a Broadly Neutralizing Nanobody Does Not Depend on Antibody Dependent Cellular Cytotoxicity.* Front Immunol, 2020. **11**: p. 627.
142. Cheung, C.Y., et al., *Development of an ELISA-Based Potency Assay for Inactivated Influenza Vaccines Using Cross-Reactive Nanobodies.* Vaccines (Basel), 2022. **10**(9).
143. Krissinel, E. and K. Henrick, *Inference of macromolecular assemblies from crystalline state.* J Mol Biol, 2007. **372**(3): p. 774-97.
144. Sok, D., et al., *The effects of somatic hypermutation on neutralization and binding in the PGT121 family of broadly neutralizing HIV antibodies.* PLoS Pathog, 2013. **9**(11): p. e1003754.
145. Muecksch, F., et al., *Affinity maturation of SARS-CoV-2 neutralizing antibodies confers potency, breadth, and resilience to viral escape mutations.* Immunity, 2021. **54**(8): p. 1853-1868 e7.
146. Wu, N.C., et al., *In vitro evolution of an influenza broadly neutralizing antibody is modulated by hemagglutinin receptor specificity.* Nat Commun, 2017. **8**: p. 15371.
147. Boder, E.T. and K.D. Wittrup, *Yeast surface display for screening combinatorial polypeptide libraries.* Nat Biotechnol, 1997. **15**(6): p. 553-7.
148. Schneider, C.A., W.S. Rasband, and K.W. Eliceiri, *NIH Image to ImageJ: 25 years of image analysis.* Nat Methods, 2012. **9**(7): p. 671-5.
149. Karlsson, R., A. Michaelsson, and L. Mattsson, *Kinetic analysis of monoclonal antibody-antigen interactions with a new biosensor based analytical system.* J Immunol Methods, 1991. **145**(1-2): p. 229-40.
150. Karlsson, R., et al., *Analyzing a kinetic titration series using affinity biosensors.* Anal Biochem, 2006. **349**(1): p. 136-47.
151. Del Rosario, J.M.M., et al., *Exploiting Pan Influenza A and Pan Influenza B Pseudotype Libraries for Efficient Vaccine Antigen Selection.* Vaccines (Basel), 2021. **9**(7): p. 741.
152. Ferrara, F. and N. Temperton, *Pseudotype Neutralization Assays: From Laboratory Bench to Data Analysis.* Methods Protoc, 2018. **1**(1): p. 8.

153. Aydin, S., *A short history, principles, and types of ELISA, and our laboratory experience with peptide/protein analyses using ELISA.* Peptides, 2015. **72**: p. 4-15.
154. Pande, J., M.M. Szewczyk, and A.K. Grover, *Phage display: concept, innovations, applications and future.* Biotechnol Adv, 2010. **28**(6): p. 849-58.
155. Pepper, L.R., et al., *A decade of yeast surface display technology: where are we now?* Comb Chem High Throughput Screen, 2008. **11**(2): p. 127-34.
156. Zahnd, C., P. Amstutz, and A. Pluckthun, *Ribosome display: selecting and evolving proteins in vitro that specifically bind to a target.* Nat Methods, 2007. **4**(3): p. 269-79.
157. Galan, A., et al., *Library-based display technologies: where do we stand?* Mol Biosyst, 2016. **12**(8): p. 2342-58.
158. Doerner, A., et al., *Therapeutic antibody engineering by high efficiency cell screening.* FEBS Lett, 2014. **588**(2): p. 278-87.
159. Hoogenboom, H.R., *Selecting and screening recombinant antibody libraries.* Nat Biotechnol, 2005. **23**(9): p. 1105-16.
160. Tsuruta, L.R., dos, M. L. , Moro, A. M., *Display Technologies for the Selection of Monoclonal Antibodies for Clinical Use.* Antibody Engineering, 2017: p. 47.
161. Dufner, P., L. Jermutus, and R.R. Minter, *Harnessing phage and ribosome display for antibody optimisation.* Trends Biotechnol, 2006. **24**(11): p. 523-9.
162. Hanes, J., et al., *Picomolar affinity antibodies from a fully synthetic naive library selected and evolved by ribosome display.* Nat Biotechnol, 2000. **18**(12): p. 1287-92.
163. Yang, G.H., et al., *Affinity maturation of an anti-hepatitis B virus PreS1 humanized antibody by phage display.* J Microbiol, 2007. **45**(6): p. 528-33.
164. Tiller, K.E., et al., *Facile Affinity Maturation of Antibody Variable Domains Using Natural Diversity Mutagenesis.* Front Immunol, 2017. **8**: p. 986.
165. Ruff, A.J., A. Dennig, and U. Schwaneberg, *To get what we aim for--progress in diversity generation methods.* FEBS J, 2013. **280**(13): p. 2961-78.
166. Delagrave, S. and D.C. Youvan, *Searching sequence space to engineer proteins: exponential ensemble mutagenesis.* Biotechnology (N Y), 1993. **11**(13): p. 1548-52.
167. Chao, G., et al., *Isolating and engineering human antibodies using yeast surface display.* Nat Protoc, 2006. **1**(2): p. 755-68.
168. Xu, M., et al., *Design and construction of small perturbation mutagenesis libraries for antibody affinity maturation using massive microchip-synthesized oligonucleotides.* J Biotechnol, 2015. **194**: p. 27-36.
169. Lim, C.C., Y.S. Choong, and T.S. Lim, *Cognizance of Molecular Methods for the Generation of Mutagenic Phage Display Antibody Libraries for Affinity Maturation.* Int J Mol Sci, 2019. **20**(8).
170. Daugherty, P.S., et al., *Quantitative analysis of the effect of the mutation frequency on the affinity maturation of single chain Fv antibodies.* Proc Natl Acad Sci U S A, 2000. **97**(5): p. 2029-34.
171. Malecek, K., et al., *Engineering improved T cell receptors using an alanine-scan guided T cell display selection system.* J Immunol Methods, 2013. **392**(1-2): p. 1-11.
172. Desmyter, A., et al., *Crystal structure of a camel single-domain VH antibody fragment in complex with lysozyme.* Nat Struct Biol, 1996. **3**(9): p. 803-11.
173. Muyldermans, S., et al., *Camelid immunoglobulins and nanobody technology.* Vet Immunol Immunopathol, 2009. **128**(1-3): p. 178-83.
174. Koenig, P., et al., *Mutational landscape of antibody variable domains reveals a switch modulating the interdomain conformational dynamics and antigen binding.* Proc Natl Acad Sci U S A, 2017. **114**(4): p. E486-E495.
175. Weiss, G.A., et al., *Rapid mapping of protein functional epitopes by combinatorial alanine scanning.* Proc Natl Acad Sci U S A, 2000. **97**(16): p. 8950-4.

176. Sidhu, S.S., *Phage display in pharmaceutical biotechnology.* Curr Opin Biotechnol, 2000. **11**(6): p. 610-6.
177. Tit-Oon, P., et al., *Prediction of the binding interface between monoclonal antibody m102.4 and Nipah attachment glycoprotein using structure-guided alanine scanning and computational docking.* Sci Rep, 2020. **10**(1): p. 18256.
178. Abbott, W.M., M.M. Damschroder, and D.C. Lowe, *Current approaches to fine mapping of antigen-antibody interactions.* Immunology, 2014. **142**(4): p. 526-35.
179. Lovell, S.C., et al., *Structure validation by Calpha geometry: phi,psi and Cbeta deviation.* Proteins, 2003. **50**(3): p. 437-50.
180. Hornak, V., et al., *Comparison of multiple Amber force fields and development of improved protein backbone parameters.* Proteins, 2006. **65**(3): p. 712-25.
181. De Genst, E., et al., *Antibody repertoire development in camelids.* Dev Comp Immunol, 2006. **30**(1-2): p. 187-98.
182. Stalin Raj, V., et al., *Chimeric camel/human heavy-chain antibodies protect against MERS-CoV infection.* Sci Adv, 2018. **4**(8): p. eaas9667.
183. Radbruch, A., et al., *Competence and competition: the challenge of becoming a long-lived plasma cell.* Nat Rev Immunol, 2006. **6**(10): p. 741-50.
184. Ellyard, J.I., et al., *Antigen-selected, immunoglobulin-secreting cells persist in human spleen and bone marrow.* Blood, 2004. **103**(10): p. 3805-12.
185. Sheng, Z., et al., *Effects of Darwinian Selection and Mutability on Rate of Broadly Neutralizing Antibody Evolution during HIV-1 Infection.* PLoS Comput Biol, 2016. **12**(5): p. e1004940.
186. Staff, P.C.B., *Correction: Effects of Darwinian Selection and Mutability on Rate of Broadly Neutralizing Antibody Evolution during HIV-1 Infection.* PLoS Comput Biol, 2019. **15**(6): p. e1007138.
187. Cobey, S., P. Wilson, and F.A.t. Matsen, *The evolution within us.* Philos Trans R Soc Lond B Biol Sci, 2015. **370**(1676): p. 20140235.
188. Rudicell, R.S., et al., *Enhanced potency of a broadly neutralizing HIV-1 antibody in vitro improves protection against lentiviral infection in vivo.* J Virol, 2014. **88**(21): p. 12669-82.
189. Zhu, J., et al., *Mining the antibodyome for HIV-1-neutralizing antibodies with next-generation sequencing and phylogenetic pairing of heavy/light chains.* Proc Natl Acad Sci U S A, 2013. **110**(16): p. 6470-5.
190. Achour, I., et al., *Tetrameric and homodimeric camelid IgGs originate from the same IgH locus.* J Immunol, 2008. **181**(3): p. 2001-9.
191. Crooks, G.E., et al., *WebLogo: a sequence logo generator.* Genome Res, 2004. **14**(6): p. 1188-90.
192. McCoy, C.O., et al., *Quantifying evolutionary constraints on B-cell affinity maturation.* Philos Trans R Soc Lond B Biol Sci, 2015. **370**(1676): p. 20140244.
193. Yaari, G., et al., *The mutation patterns in B-cell immunoglobulin receptors reflect the influence of selection acting at multiple time-scales.* Philos Trans R Soc Lond B Biol Sci, 2015. **370**(1676): p. 20140242.
194. Hsiao, Y.C., et al., *Immune repertoire mining for rapid affinity optimization of mouse monoclonal antibodies.* MAbs, 2019. **11**(4): p. 735-746.
195. Pearce, R. and Y. Zhang, *Toward the solution of the protein structure prediction problem.* J Biol Chem, 2021. **297**(1): p. 100870.
196. Moore, P.B., et al., *The protein-folding problem: Not yet solved.* Science, 2022. **375**(6580): p. 507.
197. Boder, E.T. and K.D. Wittrup, *Yeast surface display for directed evolution of protein expression, affinity, and stability.* Methods Enzymol, 2000. **328**: p. 430-44.
198. Gai, S.A. and K.D. Wittrup, *Yeast surface display for protein engineering and characterization.* Curr Opin Struct Biol, 2007. **17**(4): p. 467-73.
199. Cherf, G.M. and J.R. Cochran, *Applications of Yeast Surface Display for Protein Engineering.* Methods Mol Biol, 2015. **1319**: p. 155-75.

200. Ebrahimizadeh, W. and M. Rajabibazl, *Bacteriophage vehicles for phage display: biology, mechanism, and application.* Curr Microbiol, 2014. **69**(2): p. 109-20.
201. Boder, E.T. and K.D. Wittrup, *Optimal screening of surface-displayed polypeptide libraries.* Biotechnol Prog, 1998. **14**(1): p. 55-62.
202. Yang, W.P., et al., *CDR walking mutagenesis for the affinity maturation of a potent human anti-HIV-1 antibody into the picomolar range.* J Mol Biol, 1995. **254**(3): p. 392-403.
203. Horton, R.M., et al., *Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension.* Gene, 1989. **77**(1): p. 61-8.
204. Joska, T.M., et al., *A universal cloning method based on yeast homologous recombination that is simple, efficient, and versatile.* J Microbiol Methods, 2014. **100**: p. 46-51.
205. Swindells, M.B., et al., *abYsis: Integrated Antibody Sequence and Structure-Management, Analysis, and Prediction.* J Mol Biol, 2017. **429**(3): p. 356-364.
206. Gibson, D.G., et al., *Enzymatic assembly of DNA molecules up to several hundred kilobases.* Nat Methods, 2009. **6**(5): p. 343-5.
207. Kircher, M. and J. Kelso, *High-throughput DNA sequencing--concepts and limitations.* Bioessays, 2010. **32**(6): p. 524-36.
208. Kircher, M., P. Heyn, and J. Kelso, *Addressing challenges in the production and analysis of illumina sequencing data.* BMC Genomics, 2011. **12**: p. 382.
209. VanAntwerp, J.J. and K.D. Wittrup, *Fine affinity discrimination by yeast surface display and flow cytometry.* Biotechnol Prog, 2000. **16**(1): p. 31-7.
210. Adams, R.M., et al., *Epistasis in a Fitness Landscape Defined by Antibody-Antigen Binding Free Energy.* Cell Syst, 2019. **8**(1): p. 86-93 e3.
211. Malmqvist, M., *Surface plasmon resonance for detection and measurement of antibody-antigen affinity and kinetics.* Curr Opin Immunol, 1993. **5**(2): p. 282-6.
212. Fischer, M.J., *Amine coupling through EDC/NHS: a practical approach.* Methods Mol Biol, 2010. **627**: p. 55-73.
213. van Regenmortel, M.H. and A. Azimzadeh, *Determination of antibody affinity.* J Immunoassay, 2000. **21**(2-3): p. 211-34.
214. Neerukonda, S.N., et al., *Establishment of a well-characterized SARS-CoV-2 lentiviral pseudovirus neutralization assay using 293T cells with stable expression of ACE2 and TMPRSS2.* PLoS One, 2021. **16**(3): p. e0248348.
215. Li, Q., et al., *Current status on the development of pseudoviruses for enveloped viruses.* Rev Med Virol, 2018. **28**(1).
216. Carnell, G.W., et al., *Pseudotype-based neutralization assays for influenza: a systematic analysis.* Front Immunol, 2015. **6**: p. 161.
217. Carnell, G., et al., *An Optimized Method for the Production Using PEI, Titration and Neutralizationof SARS-CoV Spike Luciferase Pseudotypes.* Bio Protoc, 2017. **7**(16): p. e2514.
218. Van Weemen, B.K., *ELISA: highlights of the present state of the art.* J Virol Methods, 1985. **10**(4): p. 371-8.
219. Mata-Fink, J., et al., *Rapid conformational epitope mapping of anti-gp120 antibodies with a designed mutant panel displayed on yeast.* J Mol Biol, 2013. **425**(2): p. 444-56.
220. Spangler, J.B., et al., *Combination antibody treatment down-regulates epidermal growth factor receptor by inhibiting endosomal recycling.* Proc Natl Acad Sci U S A, 2010. **107**(30): p. 13252-7.
221. Oliphant, T., et al., *Antibody recognition and neutralization determinants on domains I and II of West Nile Virus envelope protein.* J Virol, 2006. **80**(24): p. 12149-59.
222. Gupta, K. and R. Varadarajan, *Insights into protein structure, stability and function from saturation mutagenesis.* Curr Opin Struct Biol, 2018. **50**: p. 117-125.

223. Rasila, T.S., M.I. Pajunen, and H. Savilahti, *Critical evaluation of random mutagenesis by error-prone polymerase chain reaction protocols, Escherichia coli mutator strain, and hydroxylamine treatment.* Anal Biochem, 2009. **388**(1): p. 71-80.
224. Krissinel, E., *Crystal contacts as nature's docking solutions.* J Comput Chem, 2010. **31**(1): p. 133-43.
225. Wrenbeck, E.E., M.S. Faber, and T.A. Whitehead, *Deep sequencing methods for protein engineering and design.* Curr Opin Struct Biol, 2017. **45**: p. 36-44.
226. Baek, M., et al., *Accurate prediction of protein structures and interactions using a three-track neural network.* Science, 2021. **373**(6557): p. 871-876.
227. Jumper, J., et al., *Highly accurate protein structure prediction with AlphaFold.* Nature, 2021. **596**(7873): p. 583-589.
228. Yan, Y., et al., *HDOCK: a web server for protein-protein and protein-DNA/RNA docking based on a hybrid strategy.* Nucleic Acids Res, 2017. **45**(W1): p. W365-W373.
229. Dominguez, C., R. Boelens, and A.M. Bonvin, *HADDOCK: a protein-protein docking approach based on biochemical or biophysical information.* J Am Chem Soc, 2003. **125**(7): p. 1731-7.
230. Eberhardt, J., et al., *AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings.* J Chem Inf Model, 2021. **61**(8): p. 3891-3898.
231. Benjin, X. and L. Ling, *Developments, applications, and prospects of cryo-electron microscopy.* Protein Sci, 2020. **29**(4): p. 872-882.
232. Danev, R., H. Yanagisawa, and M. Kikkawa, *Cryo-Electron Microscopy Methodology: Current Aspects and Future Directions.* Trends Biochem Sci, 2019. **44**(10): p. 837-848.
233. Molesti, E., et al., *Comparative serological assays for the study of h5 and h7 avian influenza viruses.* Influenza Res Treat, 2013. **2013**: p. 286158.
234. Hensley, S.E., et al., *Influenza A virus hemagglutinin antibody escape promotes neuraminidase antigenic variation and drug resistance.* PLoS One, 2011. **6**(2): p. e15190.
235. Chai, N., et al., *Two Escape Mechanisms of Influenza A Virus to a Broadly Neutralizing Stalk-Binding Antibody.* PLoS Pathog, 2016. **12**(6): p. e1005702.
236. Prachanronarong, K.L., et al., *Mutations in Influenza A Virus Neuraminidase and Hemagglutinin Confer Resistance against a Broadly Neutralizing Hemagglutinin Stem Antibody.* J Virol, 2019. **93**(2).
237. Hoffmann, M., et al., *The Omicron variant is highly resistant against antibody-mediated neutralization: Implications for control of the COVID-19 pandemic.* Cell, 2022. **185**(3): p. 447-456 e11.
238. Titong, A., et al., *First-in-class trispecific VHH-Fc based antibody with potent prophylactic and therapeutic efficacy against SARS-CoV-2 and variants.* Sci Rep, 2022. **12**(1): p. 4163.
239. Voronina, D.V., et al., *Cross-Reactive Fc-Fused Single-Domain Antibodies to Hemagglutinin Stem Region Protect Mice from Group 1 Influenza a Virus Infection.* Viruses, 2022. **14**(11).
240. Li, Q., et al., *Mucosal nanobody IgA as inhalable and affordable prophylactic and therapeutic treatment against SARS-CoV-2 and emerging variants.* Front Immunol, 2022. **13**: p. 995412.
241. Mahomed, S., et al., *Clinical Trials of Broadly Neutralizing Monoclonal Antibodies for Human Immunodeficiency Virus Prevention: A Review.* J Infect Dis, 2021. **223**(3): p. 370-380.
242. Ahangarzadeh, S., et al., *An update on antiviral antibody-based biopharmaceuticals.* Int Immunopharmacol, 2020. **86**: p. 106760.
243. Leivo, J., et al., *Engineering of a broad-specificity antibody: detection of eight fluoroquinolone antibiotics simultaneously.* Anal Biochem, 2011. **409**(1): p. 14-21.

244. Korpimaki, T., et al., *Engineering of a broad specificity antibody for simultaneous detection of 13 sulfonamides at the maximum residue level.* J Agric Food Chem, 2004. **52**(1): p. 40-7.