Exploring epitope-focused vaccine development: design of epitope scaffolds and nanoparticle presentation platforms, and computational prediction of conformational epitopes

Sergey Vitalyevich Menis

A dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

University of Washington

2013

Reading Committee:
Peter Byers, Chair
Tamir Gonen
Roland Strong

Program Authorized to Offer Degree:
Biochemistry

University of Washington

Abstract

Exploring epitope-focused vaccine development: design of epitope scaffolds and nanoparticle presentation platforms, and computational prediction of conformational epitopes

Sergey Vitalyevich Menis

Chair of the Supervisory Committee:
Dr. William R. Schief Jr.
Biochemistry

Human Immunodeficiency Virus (HIV) and Hepatitis C Virus (HCV) evolved a number of defense strategies to evade protective mechanisms of the immune system. Classical vaccine approaches have failed to elicit a protective response for these targets. Epitope Scaffolding is a theoretically attractive immunogen design strategy, which isolates known protective epitopes from their environments while stabilizing relevant conformations, as defined by a neutralizing antibody. In this work, we extend previous epitope scaffolding methods to scaffold a known protective HCV epitope and introduce preliminary immunization results. Additionally, in light of several mixed, protein and glycan, epitopes described for HIV, we apply the scaffolding strategy to an anti-HIV broadly neutralizing antibody – PG9. Next, using a derivative of an HIV envelope glycoprotein, we describe the development of a novel multimerization platform and illustrate potential applications. Finally, a computational protocol was developed to identify antibody accessible epitopes on flexible, glycosylated proteins.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I would like to thank my advisor Bill Schief for his continuous optimism in all of my projects. My committee advisors, Peter Byers, Tamir Gonen and Roland Strong, for their dedication, patience and fairness, who guided me throughout my doctoral work. It is difficult to imagine that I would be here without their supportive and, at times, stern comments. Thank you!

The Schief Lab has been an instrumental part of my growth as a scientist and I am grateful to each of you. Vanita Sood set me on the right path teaching me proper experimental technique. Yih-En Andrew Ban and Po-Ssu Huang have developed the bulk of the computational methods employed and extended throughout this work. Oleksandr Kalyuzhniy has always been an example for meticulous experimental setup and analysis.

Tinashe Ruwona from Mansun Law's Lab and Leopold Kong from the lab of Ian Wilson have been excellent collaborators for the HCV scaffolding work presented in Chapter 2. Skye MacPherson has expressed and purified a great deal of proteins used in Chapter 2. Oleksandr Kalyuzhniy carried out all of the SPR experiments. Joseph Jardine for allowing me to test my multimerization platform, described in Chapter 3, with his molecules and Jean-Philippe Julien, from Ian Wilson's lab, for diligently providing us with structures. Finally, Chris Carrico, from Roland Strong's group, for sharing the burden of developing the EPIMAP method, presented in Chapter 4.

I am thankful for the many friends I have met and was fortunate to keep throughout the years. There are too many of you to name! I am honored to have your warmth and thoughts!

My parents, Olga and Vitaliy, and sisters, Lena and Anna, for making this possible!

A special mention goes out to my friend, mentor and an occasional pain in the gluteus maximus, Bruno E. Correia!

CHAPTER 1: INTRODUCTION

1.1 Subunit Vaccines

Prophylactic vaccines are the key to reduction of morbidity and eradication of diseases (*1*, *2*). Adaptation of regular, population-wide immunization schedules led to the eradication of smallpox, as declared by the World Health Organization (WHO) in 1980. The list of vaccine-preventable diseases includes hepatitis A and B, influenza, measles, mumps, rubella and many others. Historically, vaccine formulations featured weakened, attenuated, as is the case for the rubella, or closely related, in the case of smallpox, viruses, to generate protective immunity. Alternatively, the use of inactivated/killed virus, exemplified by the inactivated polio vaccine (IPV), has also been successful. In certain cases, the use of whole virus poses a significant threat of reactivation and/or is unable to generate a strong protective immune response. Hepatitis B vaccine (Recombivax HB) is a hallmark "subunit" vaccine, composed of hepatitis B surface antigen (HBsAg) expressed in a recombinant strain of *Saccharomyces cerevisiae* (*3*). The subunit vaccine approach has since been applied to human papillomavirus (HPV) (*4*) as well. By isolating the viral antigen, we can focus the immune response on a single region of the virus and induce a strong protective response, without the risk of infecting a healthy patient.

1.1.1 Hepatitis C

Hepatitis C virus (HCV) is a major cause of liver cirrhosis, hepatocellular carcinoma and end-stage liver diseases worldwide (*5*). Despite available therapeutic treatments, prevalence of HCV has been on the rise, as much as 2.8% of the global population is

infected worldwide and nearly 10% in Mainland China (*6*, *7*). In the absence of a vaccine, current prevention strategies are based on safe blood transfusion practices and prevention of injection drug use. Treatment with ribavarin, a nucleoside analog, in combination with pegylated interferon is the current gold-standard of therapy (*8*). However, in the absence of therapy, 20% of the patients will spontaneously clear the infection suggesting that a prophylactic vaccine approach is realistic. Classical vaccine approaches, until recently, were not feasible, as there was no cell culture system to produce the virus particles on a large scale, and the live-attenuated viruses tended to persist in the host (*9*). The recently described recombinant HCV pseudo particle (HCVpp) system has demonstrated a cross-genotype protective response in non-human primates (NHPs) (*10*, *11*). Interestingly, the subunit vaccinology approach was further strengthened by a study published the same year, which demonstrated that a prime-boost approach using HCVpp separating E1 and E2 HCV envelope glycoproteins elicited a more potent neutralizing response than a more wild-type-like E1/E2 expressing HCVpp (*12*). A prophylactic anti-HCV vaccine seems within reach, and the current data suggests that recombinant approaches focusing responses to conserved viral protein regions are a promising strategy. The fact that some chronically infected patients make broadly neutralizing antibodies (bNAbs) directed against the E2 glycoprotein, while off-target interfering antibodies can reduce their neutralizing activity, support this strategy (*13*, *14*). A minimal, precisely focused subunit vaccine is needed to target induction of bNAbs but not their off-target interfering competitors.

1.1.2 Human Immunodeficiency Virus

In 2011, there were 1.7 million Acquired Immunodeficiency Syndrome (AIDS) related deaths worldwide. AIDS is caused by the Human Immunodeficiency Virus (HIV). 34 million people are living with HIV and 2.7 new infections are registered yearly (*15*). Presently, there is no prophylactic vaccine available for prevention of the HIV infection. Upon infection,

patients rely on retroviral therapy to control the virus. HIV, being a retrovirus, integrates into human DNA and thus elimination of HIV from an infected person is an extremely difficult challenge. Current standard therapeutic regimens provide no hope for clearing the infection, though extraordinary work with bone marrow transplantation has indicated that a cure is possible (*16*). As in the case of HCV, traditional vaccine approaches have failed on generating a broad protective response for HIV for several reasons, one of which is safety due to the risk of reversion to a pathogenic form (*17*). The subunit vaccinology approach has also been evaluated using components or best-approximation mimics of the cell-surface heterotrimer, composed of envelope glycoproteins gp120 and gp41. Monomeric recombinant gp120 (rgp120) has been evaluated in clinical trials without any appreciable protection (*18*, *19*). A recombinant native-like trimer was recently described, it remains to be seen whether it will be a successful vaccine candidate (*20*). Immunizations with earlier versions of recombinant trimers have not demonstrated a significant improvement over monomeric rgp120 (*17*, *21*). Despite this seemingly hopeless prognosis, 10-30% of HIV-1+ subjects do develop a cross-neutralizing response, approximately 2.5 or more years after infection (*22*, *23*). Hence, while it may be difficult to develop a broad antibody response targeting the HIV-1, due at least in part to features of the native spike such as: heavy glycosylation masking the underlying protein surface, high sequence variability of exposed loop epitopes, and steric occlusion of conserved epitopes, this task is achieved regularly, albeit inefficiently. In the past several years, a large number of quite potent bNAbs targeting at least four different regions on the native spike have been isolated from different cohorts. The main targets of bNAbs are the CD4 binding site, the V1V2 loop, the base of the V3 loop (called the N332 "supersite" (*24*)), and the membrane proximal external region (MPER) (*25-28*). Of particular interest are VRC01 (*29*), PGV04 (*30*), 12A12, NIH45-46 and 3BNC60 (*31*), all of which derive from the same human VH1-2*02 variable heavy gene despite being identified from unrelated subjects (*32*). Use of these and other antibodies in passive transfer experiments in NHPs has been demonstrated to afford protection and theoretical

exercises combining the best bNAbs predict neutralization coverage approaching 100% (*33*, *34*). Elucidation of these key antibodies and their respective modes of interactions, taken with demonstrated protection in passive transfer experiments, strengthens the notion that a prophylactic vaccine inducing bNAbs could protect against HIV acquisition and is in-fact possible.

1.2 Reverse Vaccinology

In light of poor neutralizing responses generated by anti-HIV vaccine candidates, reverse vaccinology was proposed, as an alternative vaccine development strategy (*35*). The novel approach suggested probing sera of infected individuals for existing neutralizing antibodies and determining their respective epitopes in order to guide future design of immunogens (Figure 1.1). This concept was further developed on a limited set of bNAb complexes: b12 (*36*), 4E10 (*37*), 2F5 (*38*).

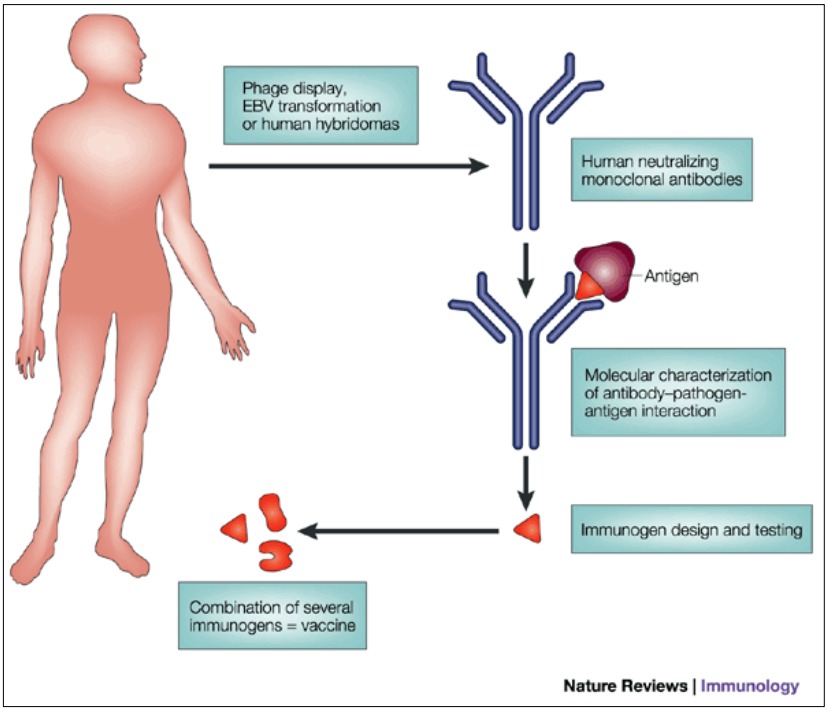


Figure 1.1 - Reverse Vaccinology. Isolation of antibodies arising from natural infection and identification of their epitopes may guide immunogen design able to relict those antibodies. Adapted by permission from Macmillan Publishers Ltd: Nat Rev Immunology (*35*), copyright 2002

Elucidation of atomic level details of multiple antibody-epitope complexes facilitates structure-based protein design of minimal immunogens specific for a given bNAb. Furthermore, as described in Section 1.1.2, numerous antibody-epitope complexes have been described recently: VRC01, PGV04, and PG9 (*39*) and others, all of which are amenable targets for future structure-based reverse vaccinology efforts.

1.3 Epitope Scaffolding

The laboratory of William R. Schief has developed a computational structure-based methodology implementing the concepts of reverse vaccinology described in the previous section. In a method called *Epitope Scaffolding*, atomic coordinates of the epitope, from the

antibody-epitope complex, deposited in the Protein Data Bank (PDB) (*40*), are used to scan the PDB in order to identify structurally similar protein backbone segments among all deposited structures, determined by X-ray crystallography. If a suitable structural match is found, the epitope is transferred onto the backbone of the matched protein scaffold. The transplanted region then undergoes several rounds of automatic design and human evaluation to identify mutations predicted to stabilize the epitope in its antibody-bound conformation within the protein scaffold. Finally, the protein scaffold carrying the epitope is examined for opportunities to improve protein stability by introducing disulfide bonds and/or circular permutation and solubility by introducing polar residues on the surface. The resulting *epitope scaffold* is a minimal, stable protein carrying the wild-type epitope, stabilized in its antibody-bound state. To date, the approach has been used to scaffold 2F5 (*41*, *42*), 4E10 (*43*, *44*) and transfer two of the three loops for the b12 HIV bNAb epitope (*45*). The methodology developed for HIV-1 bNAb epitopes is readily applicable to other systems, as illustrated by the transfer of the Motavizumab, anti-Respiratory Syncytial Virus (RSV), antibody epitope (*46*). An epitope scaffold carries the smallest possible subunit of the target virus, the bNAb epitope, while preserving the native antibody-bound conformation, which is rarely achievable by a peptide.

1.3.1 Scaffolding the HCV1 bNAb epitope

In 2012, structures of two bNAbs in complex with their epitope, targeting the same conserved region on the HCV E2 envelope glycoprotein, were determined (*5*, *47*). HCV1 is a human monoclonal antibody isolated from a humanized mouse (Medarex Inc.) immunized with soluble monomeric HCV E2 protein (*48*). AP33 is a mouse monoclonal antibody (*49*). Both isolated antibodies have broad neutralization activity across several genotypes. Their epitope, residues 412-428, is strongly conserved across clades. However, few chronically infected patients make antibodies recognizing this region (*50*). Chimpanzees are the only

model organism able to sustain an HCV infection. In these non-human primates (NHPs), passive transfer of HCV1 bNAb successfully prevented infection and therapeutic injections reduced viral load in acutely infected chimps (*51*). Finally, in chronically infected patients and vaccinated chimpanzees, interfering antibodies with epitopes overlapping or flanking the conserved 412-426 region will destructively inhibit broad cross-genotype neutralizing activity present in the sera, which is recovered upon depletion of such interfering antibodies (*14*). Immunization with HCV E2 glycoprotein can theoretically generate antibodies targeting the conserved site but they will also induce interfering antibodies. Scaffolding the HCV1 bNAb epitope will isolate the conserved region from the wild-type context and any anti-HCV antibodies induced will exclusively recognize only the 412-428 epitope, thereby avoiding induction of interfering antibodies.

1.3.2 Scaffolding the PG9 bNAb epitope

Scaffolding efforts of HIV epitopes have been carried out for several bNAbs. The HIV envelope spike is the only exposed protein component on the surface of the virion. The spike is a heterotrimeric assembly made up of three copies of non-covalently associated gp120 and gp41, anchored to the membrane by a transmembrane domain at the C-termini of gp41. All known HIV bNAbs target regions on one of the two proteins. Both gp120 and gp41 are glycosylated and present a variable sequence due to the highly error-prone HIV reverse transcriptase. Scaffolding of known bNAb epitopes removes unnecessary steric constrains of nearby glycans and selectively presents the conserved epitopes to the immune system. HIV bNAb epitopes have also served as an important platform for the development of computational methods necessary to carry out scaffolding rapidly and accurately. Over the years several linear epitopes were successfully scaffolded: 2F5 (*41*, *42*), 4E10 (*43*, *44*) HIV bNAb and Motavizumab (*46*), anti-RSV bNAb. The methods were further extended to deal with discontinuous conformational epitopes, as in the case of b12 bNAb, which was

transferred onto an unrelated protein scaffold and retained near native affinity for b12 bNAb and analysis of the b12-scaffold complex revealed near-perfect agreement with b12-gp120 complex (*45*). Given the high glycosylation level of gp120, the primary target of most known HIV bNAbs, it is inevitable that antibodies recognizing the glycan moieties will be induced. 2G12 is a unique domain exchanged bNAb, which exclusively recognizes a dense cluster of oligomannose glycans, known as the high-mannose patch (*52*). While 2G12 neutralizes clade B viruses broadly, it fails to recognize other clades as well due to variability in oligomannose distribution on those gp120s, in particular clade C gp120 (*53*). In contrast, a significant class of PGT antibodies, specifically PGT 125-128, 130 and 131, engage a mixed protein/glycan epitope and are dramatically broader, neutralizing nearly 70% of the globally circulating strains. Antibodies PG9 and PG16 recognize a different set of glycans, in addition to conserved proteins segments on the variable loops of gp120 (*54*). The structures of PG9 (*39*) and PGT128 (*53*) in complex with their respective epitopes have been recently determined. The two antibodies share a common epitope organization – protein segment sandwiched between two conserved glycans. Deletion of either of the glycans, present in the epitope, either abrogates or dramatically reduces binding to the antibody. Due to their exceptional breadth and potency PG9 and PGT128 are highly desirable targets for reelicitation and an important challenge of scaffolding a mixed protein-glycan epitope for the computational protocol.

1.4 Multivalent presentation platforms

To date, most prophylactic vaccines rely on the induction of potent T-cell dependent antibody response (*55*). Classic vaccine approaches, using whole virus, can elicit a large variety of antibodies, due the abundance of epitopes available on the virion. In contrast, subunit vaccines, especially as implemented by scaffolding, present a limited set of epitopes and can fail, if a particular epitope, presented on a scaffold, is poorly immunogenic.

Multimeric presentation platforms have been extensively used to increase immunogenicity of peptides, when chemically conjugated to a carrier particle. For protein applications, hepatitis B virus (HBV) core antigen (HBcAg) and other platforms have been used and modified extensively to display a number of different antigens by genetic fusion or chemical conjugation (*56*, *57*). Antigens can also be chemically coupled to liposomes, but they are poorly immunogenic and commonly require and adjuvant to generate T-cell help, which is critical for the induction of long-lasting memory B cells, although, some have reported that T-cell-independent B cell memory is also possible (*58*). For enveloped viruses, such as HIV and HCV, infective, non-replicating virus-like particles (VLPs) have been generated, by modifying the original virion (*10*, *59*). VLPs are a promising vaccine candidate, as they may be the most faithful mimic of the wild-type virus, still, in light of this similarity, they may elicit a similar non-neutralizing response (*59*) and in the case of HIV VLPs, immunization regiments and immunogens still need optimization, as the current responses are primarily focused on non-functional forms of HIV envelope protein on the surface (*60*). Clearly, there is a need to generate a protein-based, multivalent, recombinant antigen presentation platform, capable of presenting different antigens on the surface, with similar size and geometry as the viral antigens, which may help induce potent immune responses without the risks associated with wild-type virions (*61*). The protein-based platforms should be able to induce long-lasting B cell response in a T-cell-dependent, due to the epitopes present on the particle core and the antigen, and T-cell-independent manner, through strong B cell receptor crosslinking, due to the highly repetitive nature of the particle, if it will be indeed confirmed to be a functional mechanism, in the future.

1.5 B cell epitope prediction

Structure-based reverse vaccinology, as outlined in section 1.3 – Epitope Scaffolding, is dependent on the availability of antibody-antigen complex structures. Epitope determination

by X-ray crystallography is the most accurate experimental method to identify the epitope of a given bNAb (*62*). It is also one of the most time-intensive techniques and crystallization of complexes is often problematic. In the case of enveloped glycoproteins, the antigen may be heavily glycosylated and flexible, which makes structure determination a tremendous challenge. Ideally, using computational methods, it would be possible to accurately predict the correct structure of the antigen and the antibody from their amino acid sequence. While computational structure prediction methods have been advancing rapidly, consistent atomic-level *ab initio* prediction is only practical for protein targets, less than 100 amino acids. Still, if the structure of a viral protein is known, it should be possible to identify potential antibody epitopes computationally. A number of structure-based B cell prediction methods have been described. Most methods use some variation of solvent-accessibility measurement to rank putative epitopes. The conformational epitope predictior (CEP) is one of the first such methods, it uses an explicit solvent accessibility cutoff to identify clusters of accessible residues (*63*). DiscoTope improves on CEP by using amino acid statistics and spatial context in addition to solvent accessibility to identify prospective epitopes (*64*). An alternative method addresses the variability on the side of the antibody using a probe, from a library of 26 antibodies, with the best shape complementarity to the antigen, to be used in coarse-grained docking trials to define an antibody-antigen complex (*65*). By using a different approximation of the epitope and an improved statistic to determine side-chain exposure, Ellipro outperforms a number of other structure-based methods (*66*). Finally, PEPITO uses a combination of amino-acid propensity scores and half sphere exposure values at multiple distances, which encode solvent accessibility, to determine the epitope (*67*). Among all structure-based methods, ones that include docking perform the best and at the same time the performance of all methods is quite poor (*62*). None of these methods address the inherent flexibility of protein structures. Flexible loops can change conformation and present many different conformational epitopes. At the same time, the flexible loops can expose or hide regions of the protein in solution. Also, the constructs crystallized are

often modified to aid crystallization and flexible regions are frequently trimmed. Envelope proteins on HCV and HIV are glycosylated and this point is not addressed by any of the currently available prediction methods. A computational epitope prediction method must address protein flexibility and glycosylation, if it is going to be useful for identification of B cell epitopes present on envelope glycoproteins.

1.6 Study Overview

In this study, I will present a complete framework for design of prospective subunit vaccine candidates for reelicitation of anti-HIV and HCV bNAbs. I have also devised a general multivalent presentation platform for glycosylated immunogens. Finally, I present a high-quality computational protocol for determination of immunogenic regions on glycosylated proteins capable of exploring antigens with large flexible regions.

1) Chapter 2 will describe scaffolding of the HCV1 anti-HCV bNAb epitope (*5*). The hairpin turn conformation, observed in the crystal structure, was transplanted onto seven different protein backbones. Several variants of a single scaffold were examined to experimentally optimize the interaction with HCV1 Ab for a total of 16 variants. The scaffolds bound HCV1 with range of affinities down to picomolar range. We determined the structure of one of the scaffolds, revealing excellent agreement with our computational design. Finally, BALB/c mice were immunized with the scaffolds. After several immunizations, the sera were cross-reactive to the scaffolds and specific for the epitope. A modest level of neutralization was observed. This case study suggests that epitope scaffolding is indeed capable of producing immunogens, which can elicit neutralizing antibodies.
2) In Chapter 3, I describe my efforts to design scaffolds for the PG9 anti-HIV bNAb (*39*). This epitope includes both protein and glycan, which has not been attempted

before. Accurate transplantation of the two glycans is critical, as deletion of either eliminates or reduces affinity for the target. I designed 24 different variants based on two scaffold proteins. Over 50% of the scaffolds expressed and were purified by affinity and size exclusion chromatography. All scaffolds in one group, based on a single protein backbone, were well-behaved, thermostable monomers. Unfortunately, none of the scaffolds bound PG9 with any appreciable affinity. Given the result in Chapter 2, we are optimistic that future generations of these scaffolds will bind to PG9 with high affinity.

3) Chapter 4 is dedicated to multivalent platforms. Specifically, a 60-mer particle, based on lumazine synthase from *Aquifex aeolicus* (*68*), was used to display an engineered outer domain of HIV envelope glycoprotein gp120. A particular application of the particle is presented in context of larger work aimed at developing a vaccine candidate capable of activating calculated germline of VRC01, a potent anti-HIV bNAb (*32*). In the scope of that work, upon multimerization on nanoparticles, the designed immunogen, eOD-GT6, potently activates VRC01 germline and mature B cells. We hope to apply this platform to present the immunogens described in Chapter 2 and 3 to improve their immunogenicity.
4) Chapter 5 describes the development and multimerization of a non-germline targeted version of eOD clearly demonstrating the benefit of a multivalent presentation platform. In the absence of multimerization, eOD monomer bounds VRC01 germline with an extremely low affinity; upon display on the particle, the non-germline targeted eOD binds with high affinity and can specifically activate VRC01 germline B cells.
5) Finally, Chapter 6 presents the computational protocol for identification of immunogenic sites on flexible, glycosylated proteins, which was used in the RV144 human HIV trial analysis (*69*). The protocol employs homology modeling, to create a complete antigen, flexible loop modeling, to sample alternative protein backbone

conformations, and minimization of glycan tree conformations, to examine the effect of glycan decoration. Coarse-grained protein-protein docking, using a representative antibody probe, is applied to the ensemble of proteins generated in the previous step to produce a ranking of accessible epitopes. We hope that this protocol will be helpful to analyze the sera responses from immunization trials of HCV and HIV immunogens presented on the multimeric platforms, described in Chapter 4.

CHAPTER 2: HCV1 EPITOPE SCAFFOLDING

2.1 Introduction

Previous epitope scaffolding efforts have generated well-behaved immunogens with excellent mimicry of the antibody epitope for several HIV bNAbs (*41-45*). Despite the success of computational protein design, none of the scaffolds elicited neutralizing antibodies. In light of this, we searched for alternative systems where the feasibility of epitope scaffolding approach could be evaluated. Hepatitis C, very much like HIV, is an enveloped virus with extreme sequence variability and glycosylation of the surface antigens. Similarly, classical vaccine approaches have failed to elicit a broad protective response against HCV. Recent identification and structural characterization of two HCV bNAbs provided us with the starting points required to apply epitope scaffolding to this problem in hope of eliciting a neutralizing response.

2.1.1 Hepatitis C

Hepatitis C (HCV) is a member of the *Flaviviridae* family of enveloped viruses that features other important disease-causing agents such as the dengue virus. The virus is encoded on an ~ 11 kb strand of RNA and is translated into a large polyprotein carrying structural and non-structural proteins. The polyprotein is then cleaved post-translationally into at least 10 different proteins. The structural genes, most relevant for the scope of this manuscript, encode the capsid protein C, membrane glycoprotein prM and the envelope glycoprotein (E protein). Receptor-mediated binding and fusion occurs through the interaction with the E protein, which is the major protein on the surface of the flavivirus virion. The E protein, in the case of HCV, is composed of two proteins - E1 and E2; these form a heterodimer on the surface of the virion (*70*) (Fig. 2.1).

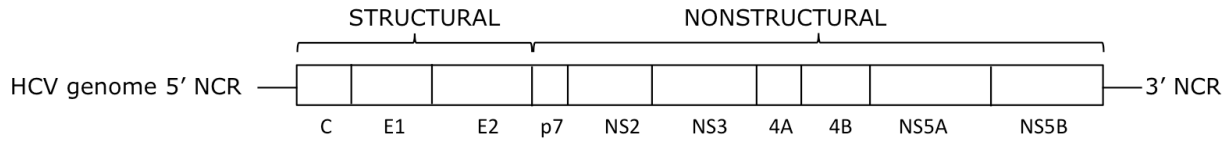


Figure 2.1 – HCV Genome Organization. Roughly 9000 base pair genome is composed of structural and non-structural segments. Envelope proteins E1 and E2 are the only proteins exposed on the surface of the HCV virion. Adapted from (*71*).

A number of molecules at the cell surface have been implicated to be important for HCV cell entry, some of which are: glycosaminoglycans, low-density lipoprotein receptor, scavenger receptor class B type I, tetraspanin CD81, tight junction protein claudin 1 and occludin (*72*). Of these, CD81 and occludin are the only two required factors to make mouse cells permissive to HCV entry (*73*).

HCV remains an important public health concern with more than 130 million people chronically infected worldwide. While approximately 25% of infected individuals will spontaneously clear the virus, the majority is at a significant risk of chronic liver disease and subsequently progressive fibrosis and cirrhosis (*74*). Presently, there is no vaccine for HCV and the current standard of care is a combination of three drugs: pegylated interferon, ribavirin, a nucleoside analog, and a protease inhibitor. To facilitate worldwide eradication of HCV, a short regiment of vaccine administration is preferred over an extended regiment of drug therapy.

2.1.2 Current HCV vaccine development strategies

Vaccines are available for several members of the Flaviviridae family such as yellow fever and promising work is under way for the dengue virus (*9*). Over the years, HCV has proven to be a much harder target for classical vaccine development approaches, primarily due to

the lack of a cell culture system to produce viral particles Hence a significant amount of vaccine design efforts were focused on recombinant protein approaches. With the advent of a pseudoparticle virus (HCVpp) (*10*) and a cell culture system (HCVcc) (*71*), the field is presently reevaluating opportunities to create a vaccine using whole virus particles. An important caveat of the HCVcc system is that it is presently limited to a single strain. The HCVpp system does not have such limitations and it is being actively evaluated as candidate vaccine with promising intermediate results (*12*).

Recombinant subunit vaccine approaches were tested early on and were shown to be safe and immunogenic, trials in healthy volunteers achieved neutralizing activity *in vitro* (*75*). Presently, several promising vaccination approaches are available and the current goal for both HCVpp-based and recombinant E1E2-based vaccines is to increase the breadth of neutralization. Zhang and colleagues have reported a cross-competition effect between neutralizing and non-neutralizing antibodies directed to the same region of the E2 protein (*14*). In their experiments, they were able to detect neutralizing activity in the plasma of chronically infected patients and chimpanzees only upon depletion of non-neutralizing antibodies directed to region flanking the neutralizing epitopes. This result suggests that while recombinant HCV glycoproteins are capable of eliciting neutralizing antibodies, off-target non-neutralizing antibodies may inhibit their activity. Isolating key neutralizing epitopes and removing any distracting regions, as described in Epitope Scaffolding (Section 1.3), is particularly suitable to address the issues described by Zhang and colleagues.

Out of the multitude of possible epitopes on both E1 and E2, humanized mouse antibodies HCV1 (*5*) and AP33 (*47*) recognize the same epitope on the HCV E2 protein and are both broad and potent. HCV1 protects chimpanzees in passive transfer experiments (*51*). Re-elicitation of such antibodies could be essential for achieving broad neutralization of HCV.

Finally, even without considering the cross-competition effect, developing a broad vaccine against HCV faces a formidable challenge of the large diversity of circulating strains, which dwarfs that of HIV (*76*). Similar to HIV, a potent and broad immune response would have to be directed to the most conserved regions of the viral glycoprotein in order to effectively neutralize a large proportion of circulating strains. The HCV1 epitope, a highly conserved patch on the HCV E2 protein, is therefore a perfect candidate, as it is conserved across a large fraction of strains.

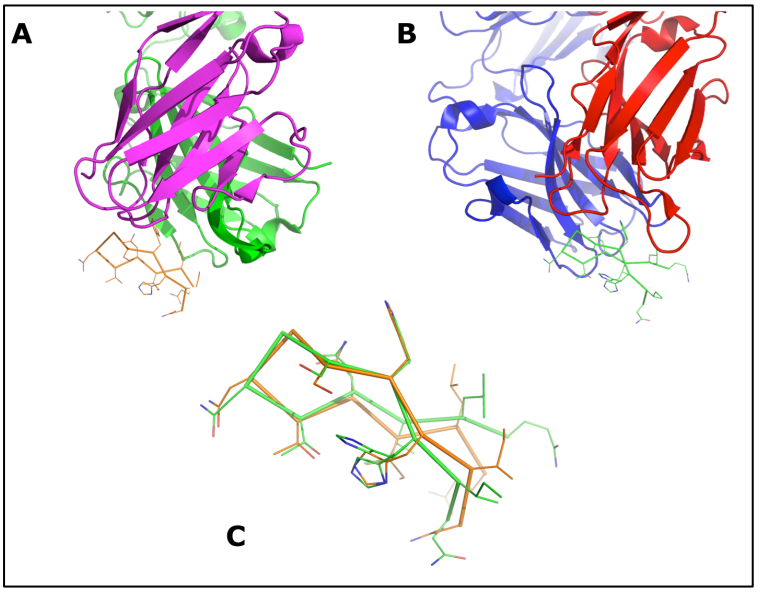


Figure 2.2 – Comparison of Ab HCV1 (PDB ID: 4DGY) and AP33 (PDB ID: 4G6A) in complex with the neutralizing epitope, residues 412-423. Both antibodies recognize the epitope in the same confirmation but their mode of binding is quite different. HCV1 in Panel A (Heavy Chain – green, Light Chain – Magenta, Peptide – orange) and AP33 in Panel B (Heavy Chain – blue, Light Chain – red, Peptide – green). Panel C: Superposition of epitopes conformations from HCV1 (orange) and AP33 (green). Figure generated in PyMOL **(*77*)**

2.2 HCV1 Epitope

The crystal structure of HCV1 broadly neutralizing antibody (bNAb) in complex with its epitope, a linear conserved region (residues 412-423) of the HCV E2 envelope glycoprotein, reveals a simple β-hairpin conformation (Figure 2.2A). The antibody primarily engages

residues 413 through 420 (*5*). This region is especially interesting, as an unrelated antibody AP33 (*47*) engages the same epitope in nearly the same conformation and is comparable in breadth and potency to HCV1 (Figure 2.2B).

HCV1 and AP33 are examples of two broadly neutralizing antibodies generated in response to immunization with recombinant E2 protein. While the binding mode for the two antibodies is quite different they both recognize nearly identical epitope conformations and engage the same key conserved side-chains (*47*).

In this work, we sought to design protein epitope scaffolds to display this neutralizing epitope in the relevant structural conformation. In an attempt to focus the immune response to the conserved epitope, we hoped to design a variety of epitope-scaffolds, using previously described methods of side-chain grafting (*41*, *43*) or backbone grafting (*42*), which would elicit potent neutralizing antibodies for HCV.

2.3 Design of HCV1 Scaffolds

2.3.1 Search and identification of protein scaffolds

The HCV1 antibody was crystallized in complex with a 12-mer (^{412}QLINTNGSWHIN^{423}) peptide (*5*). We chose to transplant only the key interacting residues (413-423). Given that the conformation is a contiguous, simple hairpin motif, we postulated that there would be a large number of high-quality protein backbone matches in the PDB (*40*). Therefore, we started by performing "superposition matching" (*43*) to identify near-perfect backbone matches for the conformation of the HCV1 epitope (^{413}LINTNGSWHIN^{423}; PDB ID: 4DGY; (*5*)) in the PDB that were compatible with the antibody binding orientation. The resulting 4249 matches were filtered down to 323 based on root mean square deviation (RMSD), between the scaffold and epitope backbones, less than 0.1 Å, scaffold size, less than 200 amino acids and clash, less than 10 Rosetta units, between the antibody and the scaffold

when aligned on the matched segment. A total of seven scaffold proteins were selected for final design (Fig. 2.3).

PDB ID:Chain	Protein	Length	Organism
3AGO:A	Ribonuclease U2	114	*U. sphaerogena*
2VBT:A	Riboflavin kinase	136	*M. jannaschii*
3HVA:A	Protein FimX	177	*P. aeruginosa*
2F6E:A	Toxin A Fragment	127	*C. difficile*
1VI4:A	Regulator of ribonuclease activity A protein 1	174	*V. cholera*
1KT9:A	Diadenosine Tetraphosphate Hydrolase	138	*C. elegans*
1WK4:B	Hypothetical protein ttk003001606	174	*T. thermophilus*

Table 2.1 – Proteins selected as potential scaffolds span a wide range of sizes and organisms, some of which are thermophilic, a beneficial quality for stable immunogens.

2.3.2 Scaffold design

In a typical side-chain grafting (*43*) protocol the relevant side-chains would be transferred to the target scaffold followed by semiautomatic design in the region of the side-chain graft. However, we have previously shown that backbone grafting, that is transfer of the backbone and side-chain atoms, can in some cases improve the results for "superposition matching" scaffolds (*42*). By transplanting the epitope backbone, it is often possible to identify stabilizing mutations underneath the epitope in the context of the scaffold. Hence, we applied the backbone grafting protocol to the seven "superposition" matches identified in the previous section. Using RosettaRemodel (*78*), we designed positions underneath the epitope to resolve steric clashes and computationally screen for any stabilizing mutations. The design stage resulted in 16 scaffold variants based on seven different protein scaffolds (Fig. 2.3 and Table 2.2).

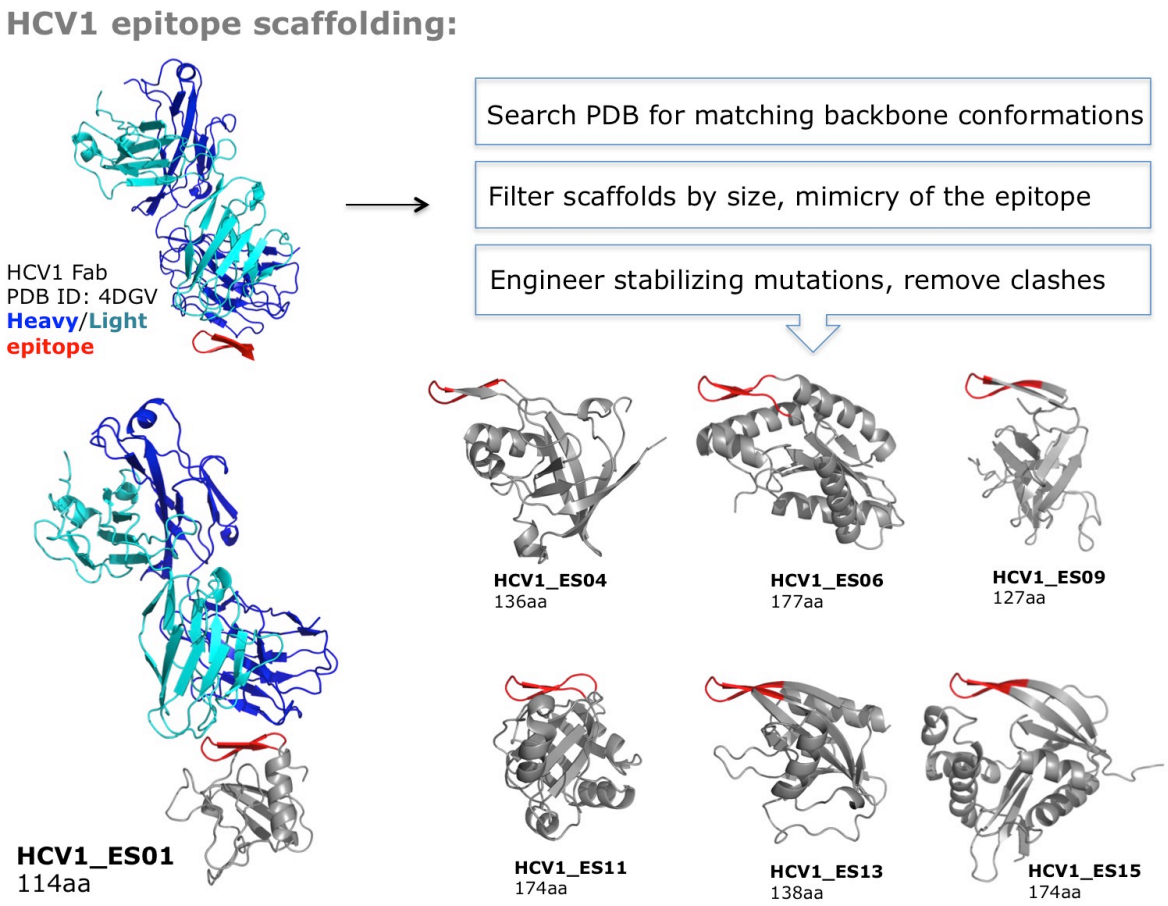


Figure 2.3 – HCV1 Epitope Scaffolding Procedure: HCV1 epitope (red) in complex with HCV1 Fab was used to search the PDB for segments with matching backbone conformations. Matches were filtered on size, mimicry of the epitope and packing underneath the epitope. Any clashes to the antibody were designed away. The resulting set of scaffolds was based on 7 different proteins (grey). Structural models were generated in PyMOL **(*77*).**

Base Protein	Scaffold Name	Additional optimizations
3AGO	HCV1_ES01	
3AGO	HCV1_ES02	I15V
2VBT	HCV1_ES03	
2VBT	HCV1_ES04	circular permutant
2VBT	HCV1_ES05	circular permutant; G85C, K88C
3HVA	HCV1_ES06	
3HVA	HCV1_ES07	L115A
3HVA	HCV1_ES08	trimmed by 30 residues
2F6E	HCV1_ES09	
2F6E	HCV1_ES10	A103G
1VI4	HCV1_ES11	
1VI4	HCV1_ES12	A129G, I131L
1KT9	HCV1_ES13	
1WK4	HCV1_ES14	
1WK4	HCV1_ES15	E162A

Table 2.2 – Scaffolds variants designed based on the seven protein scaffolds identified previously. Additional optimizations are 1) point mutations, to relieve any clashes 2) circular permutations, to improve stability and orient the epitope for multimeric fusions 3) additional disulfides and domain trims, to reduce the number of unwanted epitopes.

2.3.3 Scaffold optimization

Protein scaffolds identified by superposition matching are near-perfect structural matches for the conformation of the epitope backbone and require very little modification to accommodate the epitope side-chains. Clashes between the scaffold protein and the grafted epitope are easily resolved by the automated grafting protocol and manual design using Rosetta Remodel. Additional care must be taken to remove any clashes between the scaffold and the antibody, identify problematic residues such as unpaired cystines and remove large hydrophobic side-chains from the surface to improve protein solubility. Table 2.2 describes mutations and modifications outside of the epitope region for each scaffold. HCV1_ES02 is a variant of HCV1_ES01 with a Ile15 mutated to a valine to eliminate at small predicted clash with the HCV1 antibody. 2VBT-based scaffold, HCV1_ES03, was

competent for circular permutation by rearranging the protein sequence such that the new N- and C-termini are now at positions 85 and 86. The circular permutant of HCV1_ES03 is called HCV1_ES04 (Fig. 2.4). In our experience, circular permutation can dramatically change the stability characteristics of a protein.

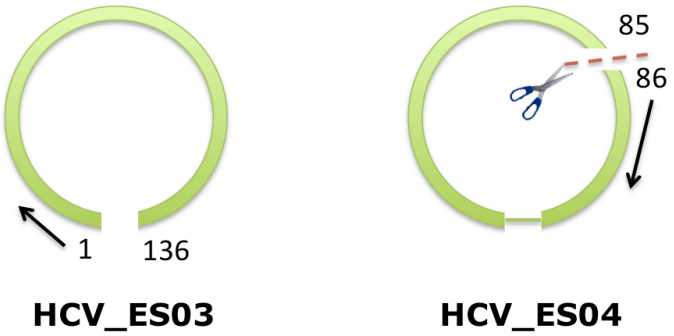


Figure 2.4 – HCV_ES04 is a circular permutant of HCV_ES03, which is a 136 amino acid scaffold based on PDBID: 2VBT. The new termini for HCV_ES04 was created by cutting between residues 85 and 86 of 2VBT. Circular permutations can help with protein stability and also orient the termini to facilitate genetic fusion to multimeric particles.

We introduced a disulfide bond to staple the new termini of HCV1_ES04, in order to further increase its stability. This construct is called HCV1_ES05. HCV1_ES07 carries a Leu to Ala mutation at position 115 to reduce potential steric clash to the antibody. Variant HCV1_ES08 lacks 30 amino acids from the N-termini of 3HVA base protein. Given a well-packed core we were confident in our ability to trim the molecule without compromising its ability to fold. HCV1_ES11 differs from HCV1_ES10 by a single Ala to Glycine mutation at position 103. Glycine is used to introduce an additional degree of flexibility to allow for the backbone to adjust to the antibody bound conformation. Finally, HCV1_ES15 is a single point (E162A) mutant of HCV1_ES14, which has a small steric clash according to Rosetta.

2.4 Biophysical characterization

Eight out of twelve scaffold variants expressed and were readily purifiable by nickel-affinity chromatography (Table 2.3). Most scaffolds were soluble except for HCV1_ES12, which was prone to aggregation. HCV1_ES09 expresses exceptionally well but also aggregates over time. In order to adequately characterize these designs, they were subject to multi-angle light scattering analysis coupled to a size-exclusion system (SEC-MALS) and their thermal stability was determined by circular dichroism (CD).

Scaffold Name	Expressed	Oligomerization state (SEC-MALS)	Expected/Observed MW (Da)	Tm (°C)
HCV1_ES01	No			
HCV1_ES02	No			
HCV1_ES03	Yes	Monomer	15403/15720	83
HCV1_ES04	Yes	Monomer	15403/14150	73
HCV1_ES05	Yes	Monomer	15307/16740	>90
HCV1_ES06	No			
HCV1_ES07	No			
HCV1_ES08	No			
HCV1_ES09	Yes	Monomer/Aggregate	14947/10630	not determined
HCV1_ES10	No			
HCV1_ES11	No			
HCV1_ES12	Yes	Aggregate	not determined	not determined
HCV1_ES13	No			
HCV1_ES14	Yes	Monomer	19199/19880	76
HCV1_ES15	Yes	Monomer	19141/20700	67

Table 2.3 – Scaffold expression, solution state, and melting temperature. Expression was determined after nickel affinity purification on a reducing SDS-PAGE gel. Oligomerization state and prediction vs. observed molecular weight was determined by SEC-MALS. Finally, the apparent thermal melting temperature was determined by monitoring protein secondary structure as a function of temperature by circular dichroism. Melting temperature for ES09 was non-determinable, as it does not have a cooperative transition.

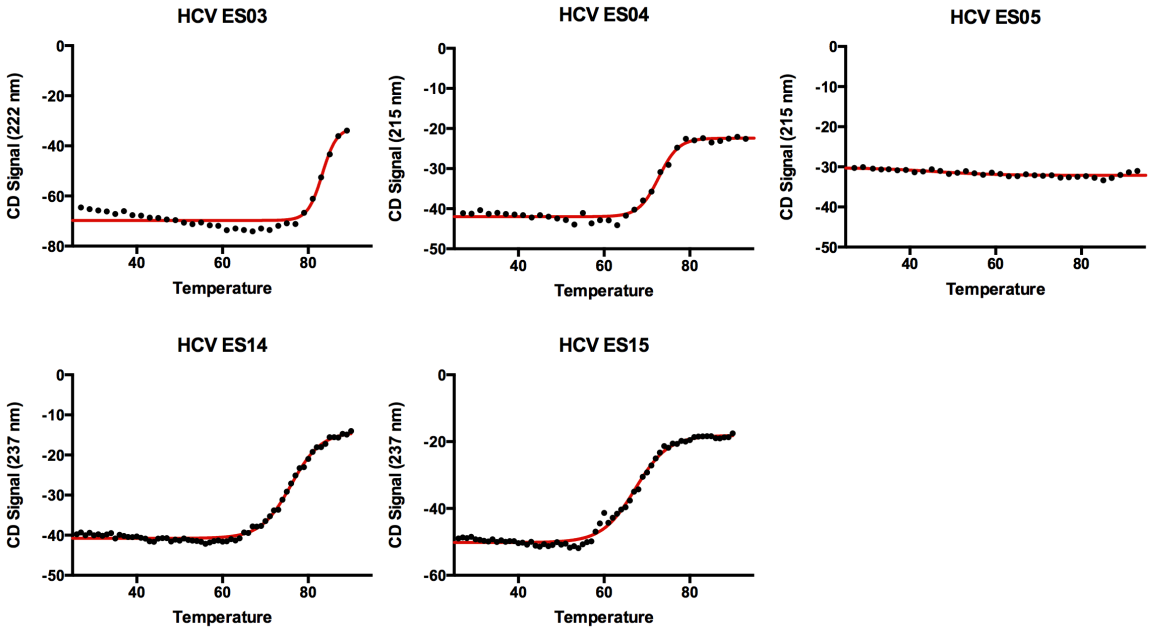


Figure 2.5 – Thermal stability analysis of HCV1 epitope scaffolds. CD signal was monitored as a function of temperature; the data was fit to a two-state model to determine the melting temperature. HCV ES04 and ES05 are circular permutations of ES03.

2.4.1 Oligomerization State and Thermal Stability

With the exception of HCV1_ES12, all scaffolds eluted in a single monodisperse peak at the predicted molecular weight, in agreement with the expected value within experimental error. It is worth noting that HCV1_ES09 also eluted in a single peak. However, there was a significant amount of aggregated material eluted as well, probably accounting for the ongoing aggregation of this molecule. We determined the thermal melting temperature by monitoring CD signal (mdeg) as a function of temperature. The curves were fitted to sigmoidal, two-state model in GraphPad Prism (Fig. 2.5). The melting temperature of ES09 could not be determined because we could not fit the data to a cooperative transition model. ES04 is a circular permutant of ES03 and is less stable. A disulfide bond on the termini of ES05 apparently prevents unfolding even at 90 degrees. HCV1_ES15 is less

stable than HCV1_ES14 due to a missing Glu (E162), which is stabilizing the fold in the wild-type protein.

2.4.2 Surface Plasmon Resonance

Monomeric and thermally stable scaffolds were tested for binding to humanized mouse HCV1 and AP33 bNAbs, as well as a panel of mouse antibodies recognizing the same epitope (Table 4.2). As the scaffolds were designed to stabilize the HCV1-bound conformation of the E2 glycoprotein epitope, affinity for HCV1 was higher than for AP33.

	Human		**Mouse**		
Scaffold Name	**HCV1 (nM)**	**AP33 (nM)**	**19B3 (nM)**	**22A9 (nM)**	**22E9 (nM)**
Peptide (LINTNGSWHI)	391±4	140±1	ND	ND	ND
HCV1_ES03	**K:** 2.455±0.002	**E:** 283±2	**K:** 3.451±0.007	**K:** 4.59±0.01	**E:** 26.8±0.1
HCV1_ES04	**K:** 1.542±0.003	**E:** 240±1	**K:** 8.03±0.03	**K:** 9.49±0.04	**E:** 65.5±0.4
HCV1_ES05	**K:** 1.094±0.002	**E:** 122±1	**K:** 5.42±0.02	**K:** 6.49±0.04	**E:** 44.5±0.03
HCV1_ES09	**K:** 1.432±0.003	**E:** 92.8±0.06	**E:** 80.9±0.3	**K:** 6.98±0.05	**E:** 24.0±0.02
HCV1_ES14	**K:** 1.362±0.002	**E:** 39.28±0.09	**K:** 0.629±0.001	**K:** 25.4±0.2	**K:** 11.05±0.07
HCV1_ES15	**K:** 0.531±0.004	**E:** 44.4±0.1	**K:** 0.445±0.001	**K:** 2.313±0.002	**K:** 1.510±0.009

Table 2.4 – Scaffold affinities determined by SPR. Values represent K_Ds in nM. A (**K**) indicates kinetic analysis and (**E**) indicates equilibrium. ND – not determined. Traces are available in the Supplementary Materials section.

All scaffolds bound HCV1 Ab with low nanomolar affinity. The affinity of HCV1_ES15 is three orders of magnitude lower than the corresponding peptide,with an apparent affinity of 531 pM. The three mouse antibodies (19B3, 22A9, 22E9) are sensitive to a range of sequence changes in the epitope region. Presently, there are no crystal structures available, but according to their binding affinities they appear to be closer to HCV1 than AP33. Oleksandr Kaluzhniy at the Scripps Research Institute performed all SPR experiments in this section.

2.4.3 X-ray Crystallography

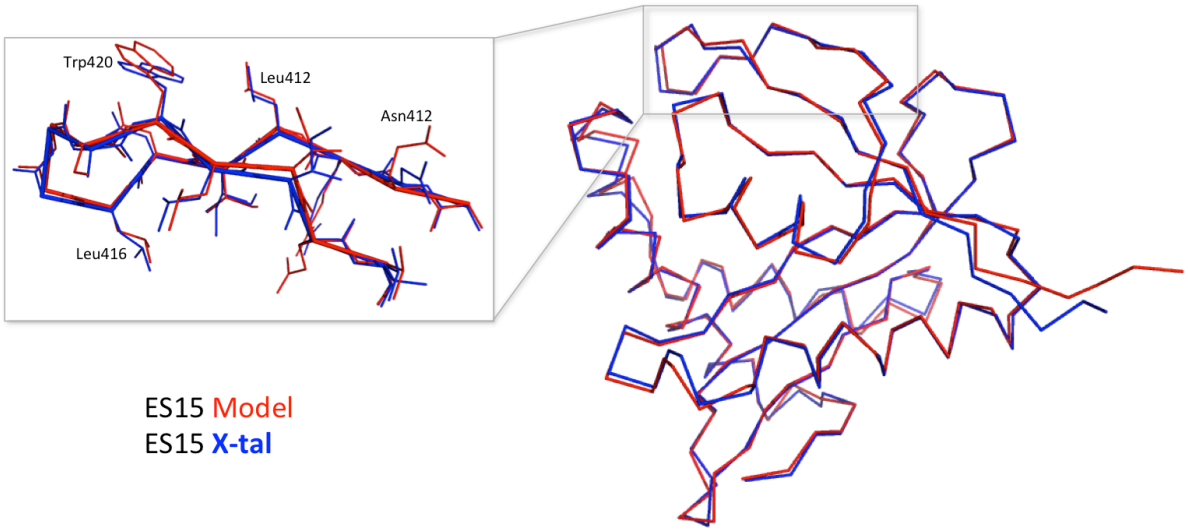


Figure 2.6 – Comparison of HCV1_ES15 model (red) to the unliganded crystal structure (blue). Key residues positions are labeled (black). The model main chain atoms are within 0.5 Å of the crystal structure. Figure generated in PyMOL **(*77*)**

To evaluate the accuracy of our computational designs, we sought to obtain high-resolution crystal structures for the scaffolds, which bound HCV1 with high-affinity. From the panel of scaffolds, only HCV1_ES09 and HCV1_ES15 formed diffraction quality crystals. The data for HCV1_ES09 was poor and we were not able to solve the structure. However, HCV1_ES15 diffracted to 1.54 Å (Table 2.5). The solved structure revealed that the crystal structure is in very close agreement with the design showing an RMSD of 0.3 Å over the main chain atoms in the epitope region (Figure 2.6). Overall, the design is within 0.5 Å of the unliganded crystal structure with a small deviation at the N-termini. This exceptional mimicry may partially explain the 531 pM affinity for HCV1 of HCV1_ES15. Leopold Kong at the Scripps Research Institute carried out x-ray crystallography and solved the structure.

Data collection	ES15 Scaffold
X-ray Source	Cu-Rotating Anode
Wavelength (Å)	1.5418
Space group	$P2_12_12_1$
Unit cell parameters	a = 44.1 b = 61.2 c = 66.2 Å
Resolution (Å)	25.15-1.85
	(1.88-1.85)[a]
Observations	108,201
Unique reflections	15,748 (571) [a]
Redundancy	6.9 (5.1) [a]
Completeness (%)	98.7 (74.3) [a]
$<I/\sigma_I>$	13.9 (7.5) [a]
R_{sym}[b]	0.10 (0.20) [a, b]
Refinement statistics	
Resolution (Å)	25.15-1.85
	(1.97-1.85)
Reflections (work)	14,909 (2392)
Reflections (test)	781 (115)
R_{cryst}(%)[c]	17.3[c] (20.0)
R_{free}(%)[d]	19.0[d] (23.2)
Average B-value (Å^2)	27.57
Wilson B-value (Å^2)	24.77
RMSD from ideal geometry	
Bond length (Å)	0.009
Bond angles (°)	1.359
Ramachandran statistics (%)[e]	
Favored	100.0
Outliers	0.0

PDB ID	---

[a] Numbers in parentheses refer to the highest resolution shell.

[b] $R_{sym} = \Sigma_{hkl}\Sigma_i \mid I_{hkl,i} - \langle I_{hkl} \rangle \mid / \Sigma_{hkl}\Sigma_i I_{hkl,I}$, where $I_{hkl,i}$ is the scaled intensity of the i^{th} measurement of relection h, k, l, $\langle I_{hkl} \rangle$ is the average intensity for that reflection, and n is the redundancy (*79*).

[c] $R_{cryst} = \Sigma_{hkl} \mid F_o - F_c \mid / \Sigma_{hkl} \mid F_o \mid \times 100$

[d] R_{free} was calculated as for R_{cryst}, but on a test set comprising 5% of the data excluded from refinement.

[e] These values were calculated using Molprobity.

Table 2.5. Data collection and refinement statistics

2.5 Immunization of BALB/c mice

HCV1 and AP33 are both humanized antibodies from mice immunized with purified E2 HCV glycoprotein. Given the close mimicry of the scaffolds of the HCV E2 segment and the corresponding high-affinities of the scaffolds for both HCV1 and AP33 Ab, we wanted to test the ability of the scaffolds to reelicit such antibodies by immunizing BALB/c mice. We chose to immunize four groups of mice: one group per scaffold (HCV1_ES05, HCV1_ES09, HCV1_ES15) and a separate group immunized with a mix of scaffolds. The mice were immunized at regular intervals and sera was collected after three injections (Figure 2.9). Tinashe Ruwona at the Scripps Research Institute carried out all immunizations and subsequent analysis.

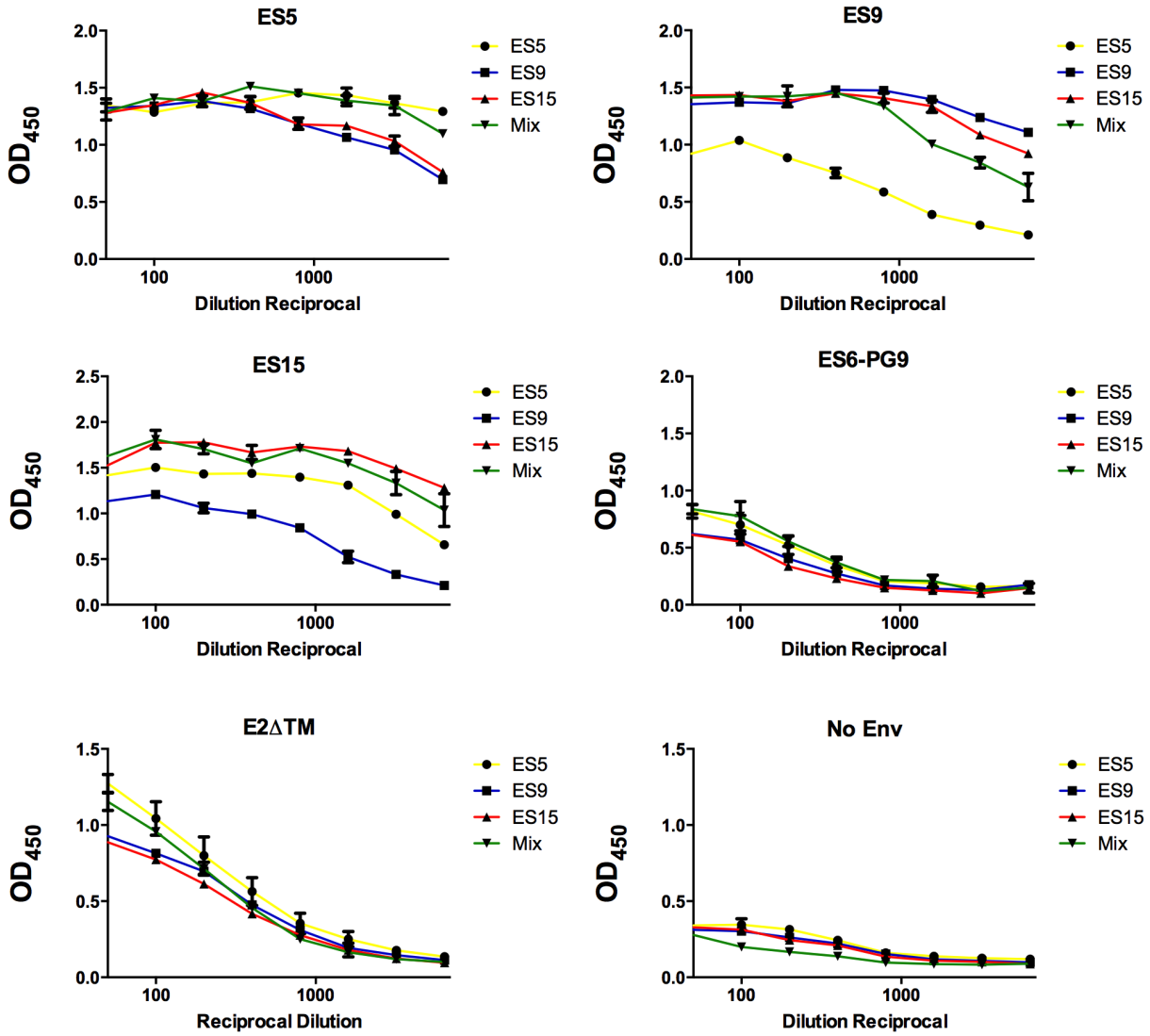


Figure 2.7A – Antibody titers of pooled mouse sera at Bleed 1, after 3 immunizations with scaffold. Sera were serially diluted from 1:50 to 1:6400 and their reactivity against immobilized scaffolds (ES5, ES9, ES15, scaffold mix and soluble E2 protein) was determined by ELISA. ES6 PG9 and No Env were negative controls. Sera from all animal groups were cross-reactive to the epitope and the wild-type protein.

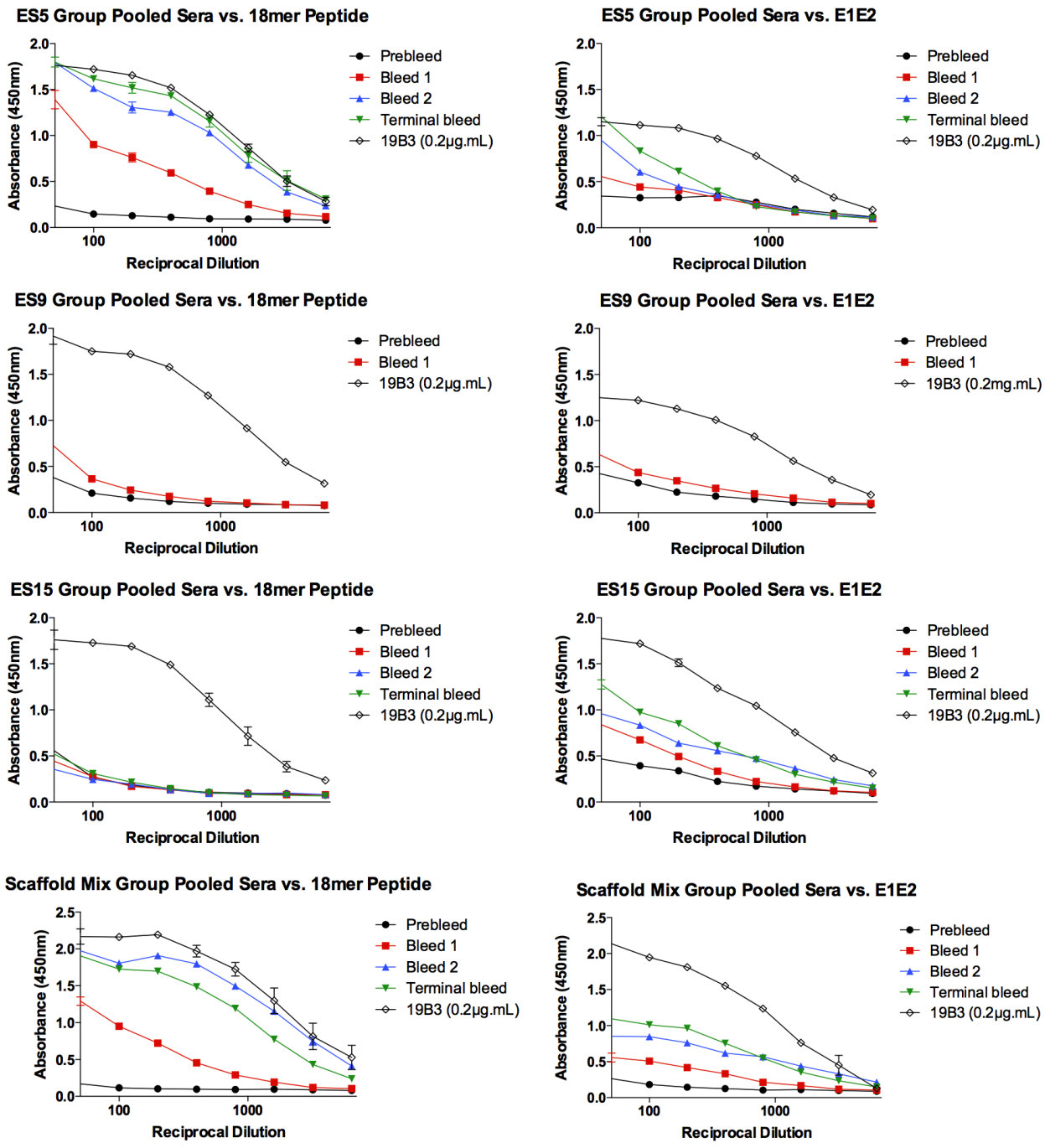


Figure 2.7B – Sera cross-reactivity for all time points measured to HCV1 18mer epitope peptide and E1E2 cell lysate. Data past Bleed 1 for ES9 immunized group is unavailable, as a significant number of animals died. Polyclonal sera from group ES5 achieves the same level of binding for 18mer peptide, as the monoclonal 19B3 antibody, used as a positive control. ES Mix combines the promising reactivity of ES5 for the peptide and E15 for E1E2 lysate.

2.5.1 Scaffold Cross Reactivity

After three immunizations sera from all animal groups was cross-reactive to the shared scaffolded epitope (Figure 2.7A). Sera from all groups of animals potently bound ES5. Animals immunized exclusively with ES5 did not bind HCV1_ES9 captured on the ELISA plate. Similarly, ES9 immunized animals had a reduced affinity for ES15 in ELISA as compared to other animal groups. ES6-PG9 was included as a negative control molecule from an unrelated experiment. None of the sera bound to the negative control with any appreciable affinity, excluding the possibility of cross-reactivity through the poly-histadine tag.

Sera from all groups recognized HCV1 18mer peptide (Figure 2.7B). Notably, ES15 and ES9 immunized groups demonstrated lower response to the peptide, as opposed to ES05 group, which may be a consequence of elevated rigidity of the epitope on the scaffold and the antibodies specifically tuned to that conformation. A peptide from IGH526 was a negative control.

When tested against purified E2 with the transmembrane region deleted, all immunized groups responded adequately (Fig 2.7B). This result suggests that all immunized groups make antibodies specific to the epitope, as demonstrated by cross-reactivity between groups of animals to all scaffolds. All sera recognize the purified E2. Hence, all animals produce antibodies toward the HCV1 epitope and the antibodies are biologically relevant, as they recognize the epitope in the context of the native protein.

It is important to note that two boost injections of purified E2 protein, between Bleed 2 and Terminal Bleed, did not improve antibody titers significantly over the scaffold-only levels.

		VSV			H77			J6E3			UKN1B12		
	Bleed	1	2	Terminal	1	2	Terminal	1	2	Terminal	1	2	Terminal
ES5	M12116	17	7	13	47	25	51	17	23	29	22	11	13
	M12117	18	-4	10	44	41	34	16	27	20	33	24	19
	M12118	-2	3	22	20	31	37	18	26	30	0	7	24
	M12119	3	3	10	29	35	59	7	16	40	17	14	18
	M12120	-9	1	3	37	44	42	3	23	20	7	5	23
ES9	M12121	-	-	-	-	-	-	-	-	-	-	-	-
	M12122	25	-	-	45	-	-	22	-	-	34	-	-
	M12123	30	-	-	37	-	-	20	-	-	39	-	-
	M12124	33	-	-	33	-	-	49	-	-	54	-	-
	M12125	27	-	-	48	-	-	41	-	-	28	-	-
ES15	M12126	26	8	-	22	38	-	40	33	-	41	22	-
	M12127	5	-30	2	30	28	38	22	13	24	30	40	37
	M12128	-26	11	10	37	43	34	-11	32	21	-13	22	29
	M12129	12	19	4	34	35	39	1	40	29	38	26	30
	M12130	2	27	-	40	42	-	1	34	-	7	3	-
ES Mix	M12131	-	-	-	-	-	-	-	-	-	-	-	-
	M12132	19	15	16	41	43	49	35	48	41	21	23	17
	M12133	24	-19	20	34	50	49	36	48	45	33	36	15
	M12134	-1	26	13	42	51	52	36	23	24	17	14	24
	M12135	8	30	17	44	45	47	13	12	13	7	9	22
Prebleed		14	3		23	18		15	18		-6	34	
		7	-4		8	12		11	9		16	12	
Control (AR3A, 20ug/ml)		6	-	-	95	-	-	74	-	-	77	-	-
Control (19B3, 20ug/ml)		-	14		-	78		-	64		-	69	

%Neutralization	<25	25-50	>50

Figure 2.8 – Neutralization data across three time points. Vesicular stomatitis virus (VSV) is a negative control for non-specific neutralization. AR3A (human) and 19B3 are positive control neutralizing antibodies. All groups modestly neutralized H77. ES5 and ES15 demonstrate low but significant (Figures S7 and S8) neutralization of J6E3. Sera from ES15 group also neutralized UKN1B12.

2.5.2 Neutralization activity

All groups of animals displayed modest neutralization across several strains, as tested in a pseudovirus infectivity reduction assays (Fig. 2.8). Pseudotype virus particles displaying vesicular stomatitis virus envelope glycoprotein G (VSVpp) were used as a control for non-specific neutralizing activity. Only one group, ES9, had high background signal for VSPpp.

Sera from the terminal bleed for all groups robustly neutralized the H77 strain of HCV. Some animals in the ES5 immunized group had neutralization levels (59%) comparable to that of the monoclonal mouse 19B3 antibody (78% for H77 at 20 ug/ml). Modest but statistically significant levels of neutralization were observed for mice immunized with ES5

for H77 and J6E3 strains; for mice immunized with ES15 for H77, J6E3 and UKN1B12; for mice immunized with a mix of three scaffolds for the H77 strain (Figure S7 and S8).

It is unclear whether E2 boost injections had any effect on neutralization, as there was very little change in neutralization potency observed between Bleed 2 and Terminal Bleed time points.

This result suggests that HCV1 epitope scaffolds elicit neutralizing antibodies specific for the HCV1/AP33 epitope. Their breadth and potency remains an open target for optimization.

2.6 Methods

2.6.1 Scaffold selection and design

The scaffolds were identified by superposition matching, as described previously (*42*). HCV E2 residues (^{413}LINTNGSWHIN^{423}) were matched against a curated subset of high-resolution PDB structures, containing 26,631 monomeric protein chains. The resulting set was filtered for potential protein scaffolds, which are nearly clash free, less than 50 Rosetta units of intra- and inter-protein clash, small, less than 200 residues and are near-perfect backbone matches for the epitope conformation, as observed in the HCV1 antibody complex structure (PDB ID: 4DGY).

Matches identified by side-chain grafting were aligned to the target scaffold protein replacing the native backbone in the matched region. RosettaRemodel (*78*) was used to close the chain breaks and introduce stabilizing mutations underneath the epitope.

2.6.2 Expression and Purification

Scaffolds were synthesized in a pET-29b(+) vector (GenScript), transformed into Arctic Express E. Coli cells (Stratagene) and plated on kanamycin LB plates. Single colony was used to inoculate a 100 mL LB (with kanamycin at 1 mg/mL) starter culture and grown overnight at 37 °C. 10 mL of the starter culture was expanded into 1L of LB with kanamycin (1 mg/ml) and grown to absorbance of ~0.5 and 220 rpm at 37 °C. The 1L culture was cooled to 12 °C and induced with 500 µL of IPTG (Sigma-Aldrich). Cells were harvested using a SLC-4000 rotor (Beckman) for 15 minutes at 6000 rpm and resuspended in start buffer (160 mM imidazole, 4 M NaCl, and 20 mM Na_2PHO_4, pH 7.4) with one tablet of ethylenediaminetetraacetic acid (EDTA)-free Protease Inhibitor (Roche), and frozen at −20 °C until ready for processing. 10 mL of 10× Bugbuster (Novagen), 20 µL of Benzoase Nuclease (Novagen), and 1.7 µL of rLysozyme (Novagen) were added to the resuspended cells, which were then lysed by rocking at room temperature for 20 min. Filtered supernatant was loaded onto a 5 mL HisTrap FF (GE Healthcare) and eluted with 500 mM imidazole, 500 mM NaCl, and 20 mM sodium phosphate, pH 7.4. Relevant fractions were concentrated in Amicon Ultra-15 (Millipore). Proteins were further purified using size-exclusion chromatography on a Superdex 75 16/600 column (GE Healthcare) in Hepes-buffered saline (HBS) buffer (10 mM Hepes, 150 mM NaCl, and 3 mM EDTA, pH 7.4). The collected fractions were analyzed on a 4–12% SDS denaturing gel (Invitrogen). Positive fractions were combined, and the final concentration was determined by measuring the UV absorption signal at 280 nm (Nanodrop 1000).

2.6.3 Light Scattering

The oligomerization state of the proteins in solution was determined by static light scattering (miniDAWN TREOS, Wyatt) coupled in-line to HPLC (Agilent, 1200 series) and a refractometer (Optilab® T-rEX™, Wyatt). Prior to entering the light scattering module the sample was further separated by Superdex 75 10/300 (GE Healthcare). One hundred twenty

microliters of each protein sample at 1–3 mg/mL were run in phosphate-buffered saline and the resulting data was analyzed using ASTRA software (Wyatt).

2.6.4 SPR

All binding experiments were carried out on a Biacore 2000 instrument (GE Healthcare) at 25 °C with HBSEP+BSA (0.01 M HEPES pH 7.4, 0.15 M sodium chloride, 3 mM EDTA and 0.005% (v/v) Surfactant P20) (GE Healthcare) + 1mg/mL BSA as running buffer. For binding analysis, ~300 response units (RUs) of IgG were captured on a CM5 sensor chip containing 8000-9000 RUs of amine-linked mouse anti-human IgG (Human Ab Capture kit, GE Healthcare). Samples of different protein concentrations were injected in duplicates over this surface at a flow rate of 50 µL/min. If necessary, surface regeneration was performed with two 60 s injections of 3 M magnesium chloride at a flow rate of 10µL/min. One flow cell contained anti- human IgG only and its interaction with the analyte was used as reference. Data preparation and analysis were performed using Scrubber 2.0 (BioLogic Software). For kinetic analysis, biosensor data were globally fit to a mass transport limited simple bimolecular binding model.

2.6.5 X-ray Crystallography

Unliganded HCV1_ES15 scaffold crystallized over a period of 28 days at 20 °C in a crystallization reagent consisting of 20% (w/v) PEG 3350 and 0.2 M Ammonium Iodide (condition from JCSGI E4). Crystals were harvested and cryoprotected by a brief immersion in 70% well buffer, 30% glycerol, followed by immediate flash-cooling in liquid nitrogen. Data were collected in house with a MicroMax™-007 HF microfocus X-ray generator equipped with VariMax HF optics and a MAR345DTB image plate using a copper rotating anode (wavelength: 1.5418 Å) cooled at around 100 K (MSC Xstream). Data were

processed and scaled with HKL-2000 (*80*). All data processing statistics are summarized in Table 2.5.

The unliganded HCV1_ES15 scaffold structure was determined by the molecular replacement method using Phaser with the computational model derived from ROSETTA. Model building was carried out using Coot-0.7 and refinement was implemented with the Phenix program (*81*). See Table 2.5 for final refinement statistics.

2.6.6 Immunizations

Female specific-pathogen-free inbred BALB/c mice were purchased from Scripps Research Institute (TSRI) Breeding Colony at 6 to 8 weeks of age. Animals were housed in ventilated microisolator cages under environmentally controlled conditions at the TSRI animal facility in compliance with AAALAC guidelines and an approved IACUC protocol. Four groups of 5 mice each were immunized with 50 µg/animal of scaffold protein (ES05, ES09, ES15 or a mix of three) in (R848) Imidazoquinoline + Monophosphorly Lipid A (MPLA) (TLR7/8 + TLR4) in 50 % Alum (Androgel 2% Aluminim Hydroxide gel) (Invitrogen) per protocol described in Figure 2.9. Animals were subsequently boosted with a purified truncated E2 glycoprotein (isolate H77 residues 412-645) at 10 µg/animal administered intraperitoneally without adjuvant twice and terminal bleeds under anesthesia were collected 14 days after last boost.

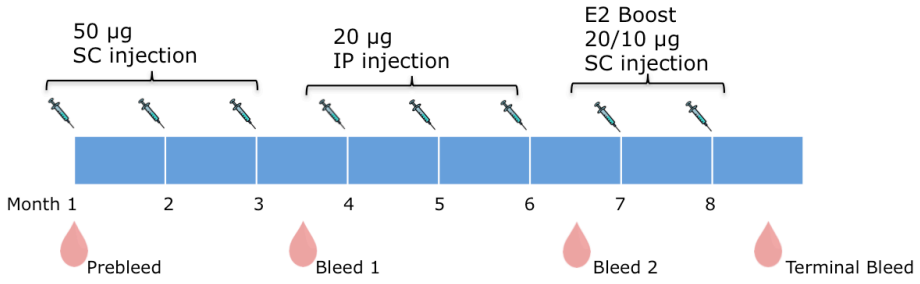


Figure 2.9 – Immunization protocol. The animals were immunized three times subcutaneously (SC) followed by three intraperitoneal injections (IP). Each animal was subsequently boosted with recombinant E2 protein. Bleeds were taken after each set of immunizations.

2.6.7 ELISA Cross-Reactivity

Antibody titers of pooled mouse sera to A) immunizing antigens (ES5, ES9 and ES15), B) HCV1 18mer peptide (AKQNIQLINTNGSWHINS) and E1E2 and E2 cell lysates. Sera were serially diluted from 1:50 to 1:6400 and their reactivity, against E1E2 cell lysate, was determined by ELISA. ES6 PG9, IGH526 peptide and No Env were negative controls.

ELISA plates were coated with 5 μg/ml of epitope scaffold or peptide and serially diluted sera incubated for 1 hour. Binding of antibodies was detected using HRP-conjugated goat anti-mouse (Ig(G+M)) F(ab')2 antibody (1:2000) (Jackson Laboratories) and TMB substrate. For part C, E1E2 and E2 antigens were captured onto ELISA wells precoated with Galanthus nivalis lectin.

2.6.8 ELISA Neutralization

HCV pseudotype virus (HCVpp) was generated by co-transfection of 293T cells with pNL4-3.lucR−E− and the corresponding expression plasmids encoding the E1E2 genes (H77, U.K.N1b12.6, J6E3) at 4:1 ratio by polyethylenimine. Virus infectivity was detected with the firefly luciferase assay system (Promega), and percentage neutralization was calculated as residual virus infectivity at the indicated antibody concentrations divided by infectivity without antibody after background subtraction. Background infectivity of the pseudotype virus was defined by infecting cells with virus made with pNL4-3.lucR−E− only. Pseudotype virus particles displaying the vesicular stomatitis virus envelope glycoprotein G (VSVpp) were a control for nonspecific neutralizing activity. The virus was incubated with the diluted sera/antibodies (1:50) for 1 h at 37°C before adding to Huh-7 cell monolayers and incubation for another 6 hours. Media was changed and measurements done 72 hours later.

2.6.9 Circular Dichroism

The stability of the proteins was measured by circular dichroism on an Aviv 420 spectrometer by far-UV (200–260 nm) wavelength scan followed by thermal melting. Two hundred fifty microliters of each protein, at a concentration between 10 and 50 μM, in phosphate-buffered saline, was analyzed in a 1-mm cuvette. Experiments were carried over a temperature range from 25 to 95°C, with 1 or 2°C increments every 3 min, and fitted to a two-state model.

2.7 Discussion

The work presented in this manuscript bridges the gap between the epitope scaffolding work in the HCV and HIV fields (*42*, *43*, *45*, *46*). The protein design methods developed to deal with HIV epitopes can be applied to protective epitopes for which a crystal structure is available. The exceptional agreement of the X-ray crystal structure and the design model highlights the robustness of the method (Figure 2.6 and Table 2.5).

In the current immunization protocol, we chose to evaluate the highest achievable level of neutralization using single monomeric scaffolds. All of the animal groups were cross-reactive even with a single scaffold immunization. The same is true for the animals immunized with a mix of scaffolds concurrently (Figure 2.6). The scaffolds provide a minimal version of the desired protective epitope but they are deficient of any native constraints present on the E2 protein and/or E1E2 dimer. Therefore, we boosted the animals with purified E2 protein to further select for antibodies, which are relevant to the native context. It is fair to note that E1E2 dimer, in theory, provides the most biological context and could be a more suitable boost. Given the choice of the available scaffolds, E1, E2, and E1E2 dimer, as well as HCVpp, the composition and timing of the immunization protocol is a major variable, which should be tested extensively.

Despite recognizing the same epitope in nearly identical conformation, HCV1 (*48*) was isolated from a HuMab-Mouse (Genmab: transgenic mouse producing human antibodies) and AP33 (*82*) was identified from a BALB/c mouse. These two antibodies have shown that given the right antigen, it is possible to elicit potent neutralizing response to the same key epitope in two different systems, with an antibody repertoire very different from each other. In theory, the gold standard for our scaffolds would be reelicitation of these antibodies in their respective systems. However, given the different immunogenic response of the scaffolds as opposed to the E2 protein, it is prudent to test the antigen in different systems, as modest polyclonal sera neutralization in one system (BALB/c mouse) may not exclude the possibility of potent neutralization in another. It is important to test these immunogens in HuMab-Mouse system and ideally chimpanzees. Finally, future work should include isolation of monoclonal antibodies and their characterization to explore possible modifications to improve their potency.

A design strategy, presented in Chapter 3, is multimerization. The HCV virion is typically between 30 to 50 nm, depending on glycosylation (*70*). The multivalent particle system presented in Chapter 3 displays 60 copies of an antigen and has a 30 nm diameter, which is nearly identical to that of the HCV virion. Future work will focus on presenting the HCV scaffolds described in this manuscript or a second generation of scaffolds on these particles. By presenting the scaffolds on the particles, we should be able to get a significant increase in overall antibody response and potentially undergo the same processing the HCV virion in the course of natural infection.

Overall, the level of neutralization achieved by the described monomeric scaffolds has not been observed in any of the previous immunization trials of epitope scaffolds. We are now confident that epitope scaffolding is a practical vaccine design strategy, which can elicit neutralizing response towards a specific epitope.

2.7 Supplementary Materials

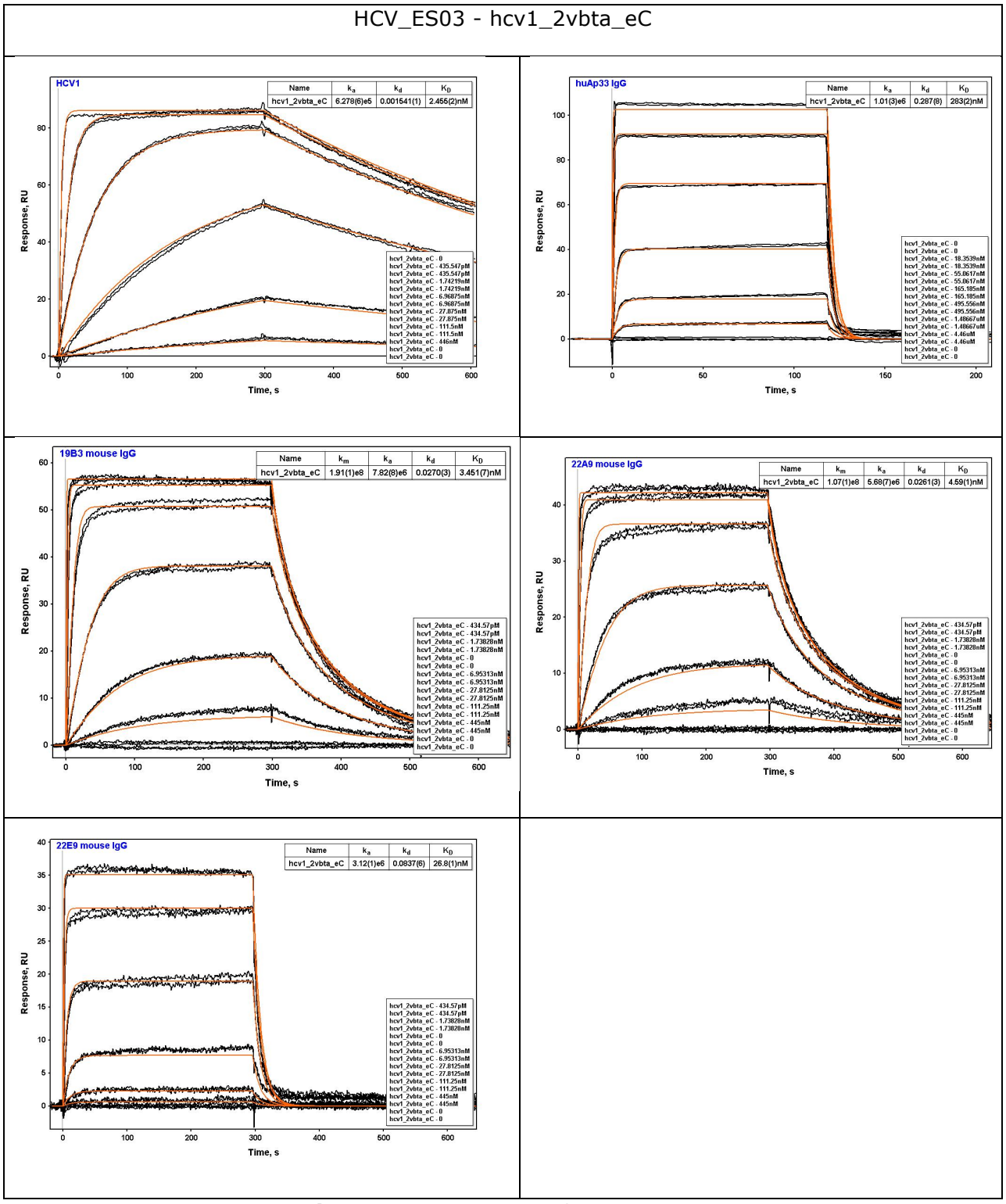


Figure S1 – SPR traces for HCV ES03

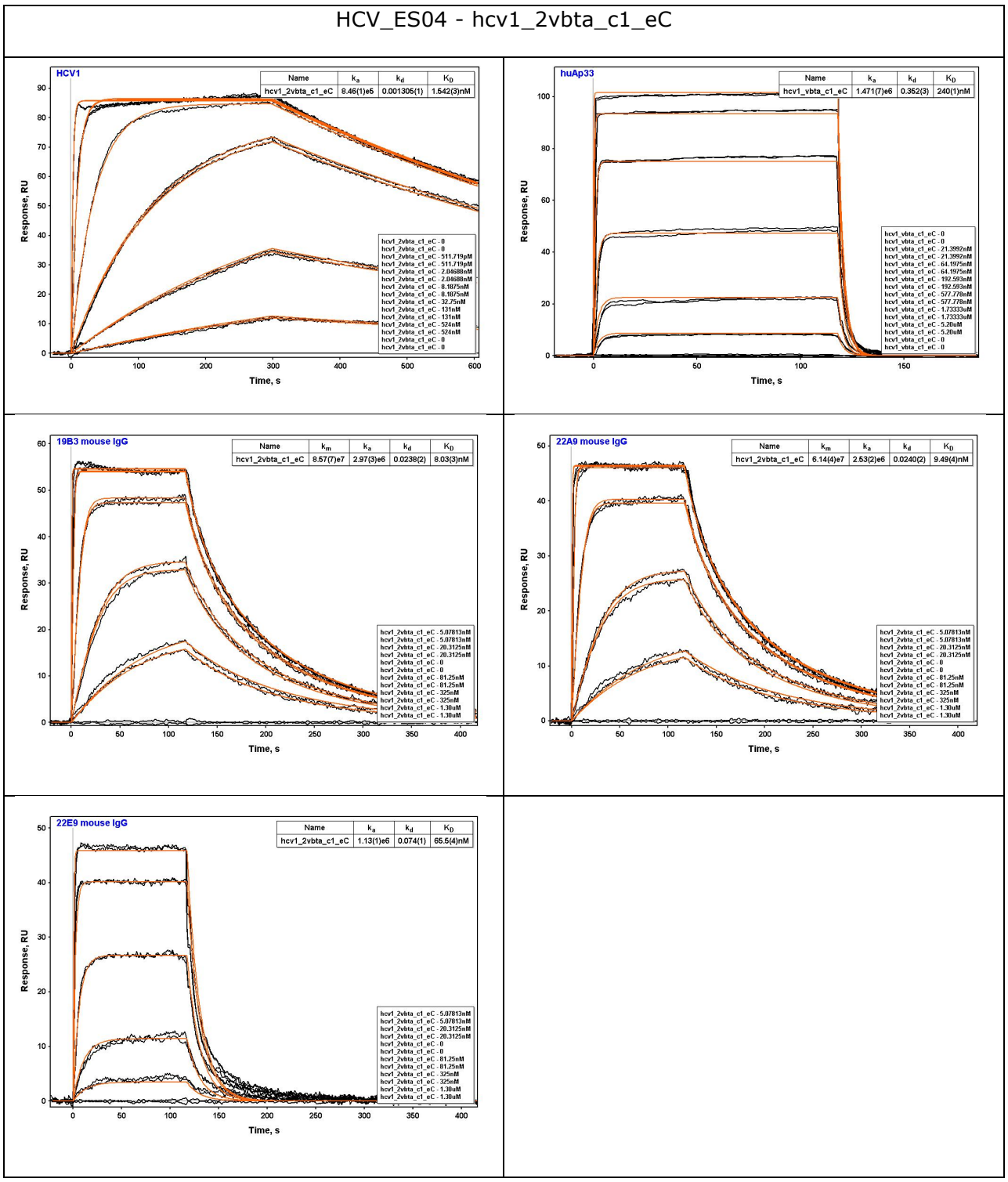


Figure S2 – SPR traces for HCV ES04

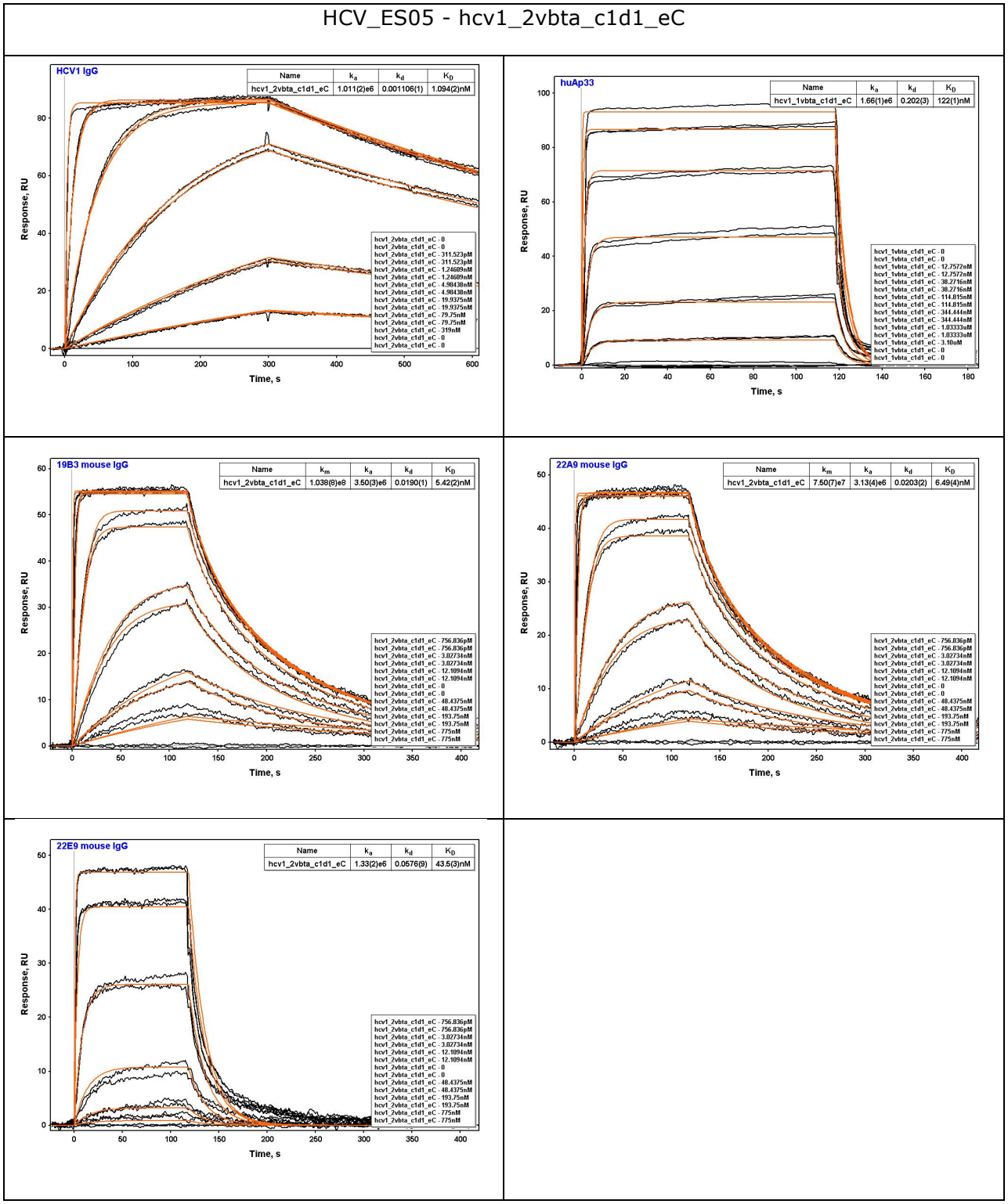


Figure S3 – SPR traces for HCV ES05

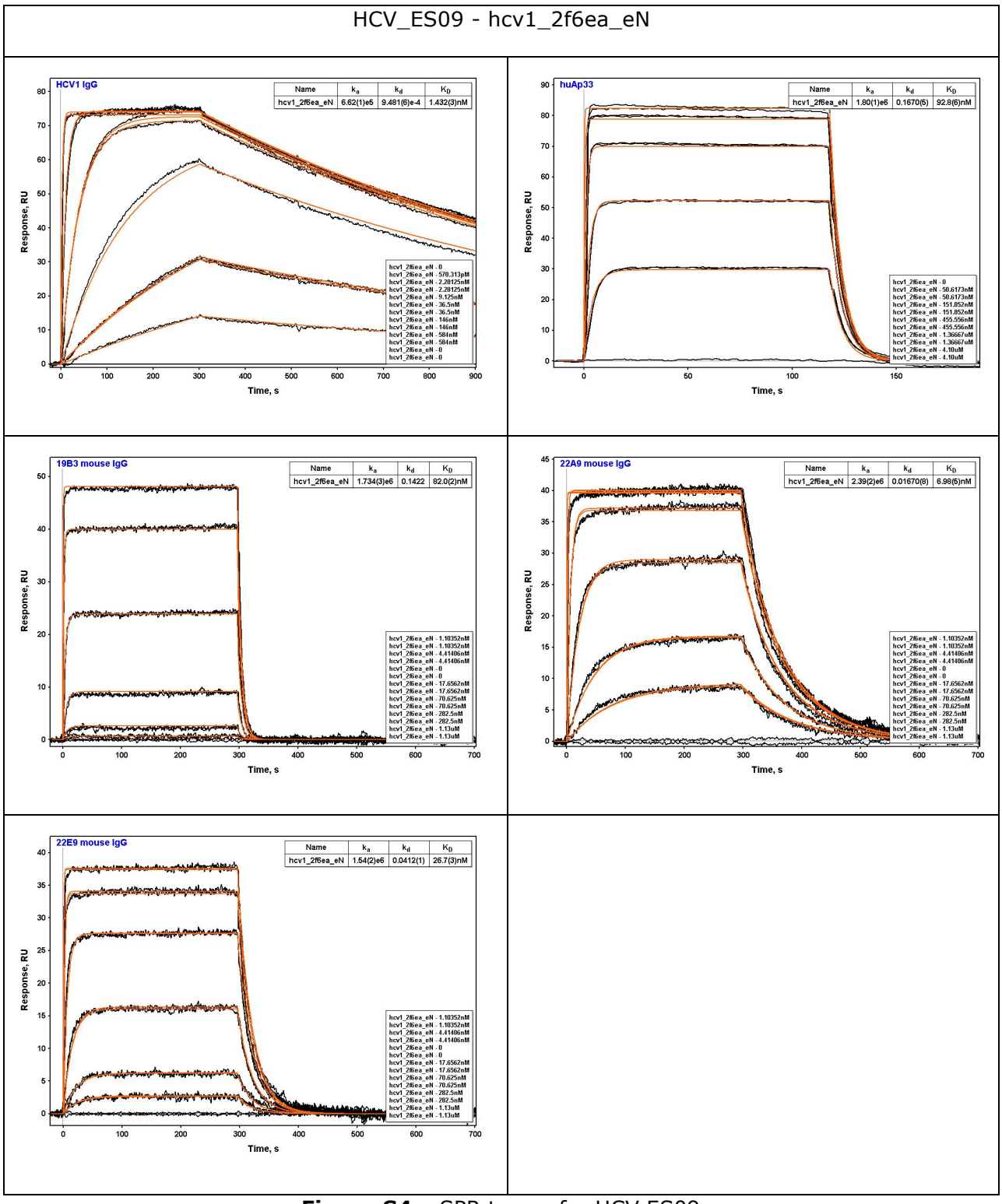


Figure S4 – SPR traces for HCV ES09

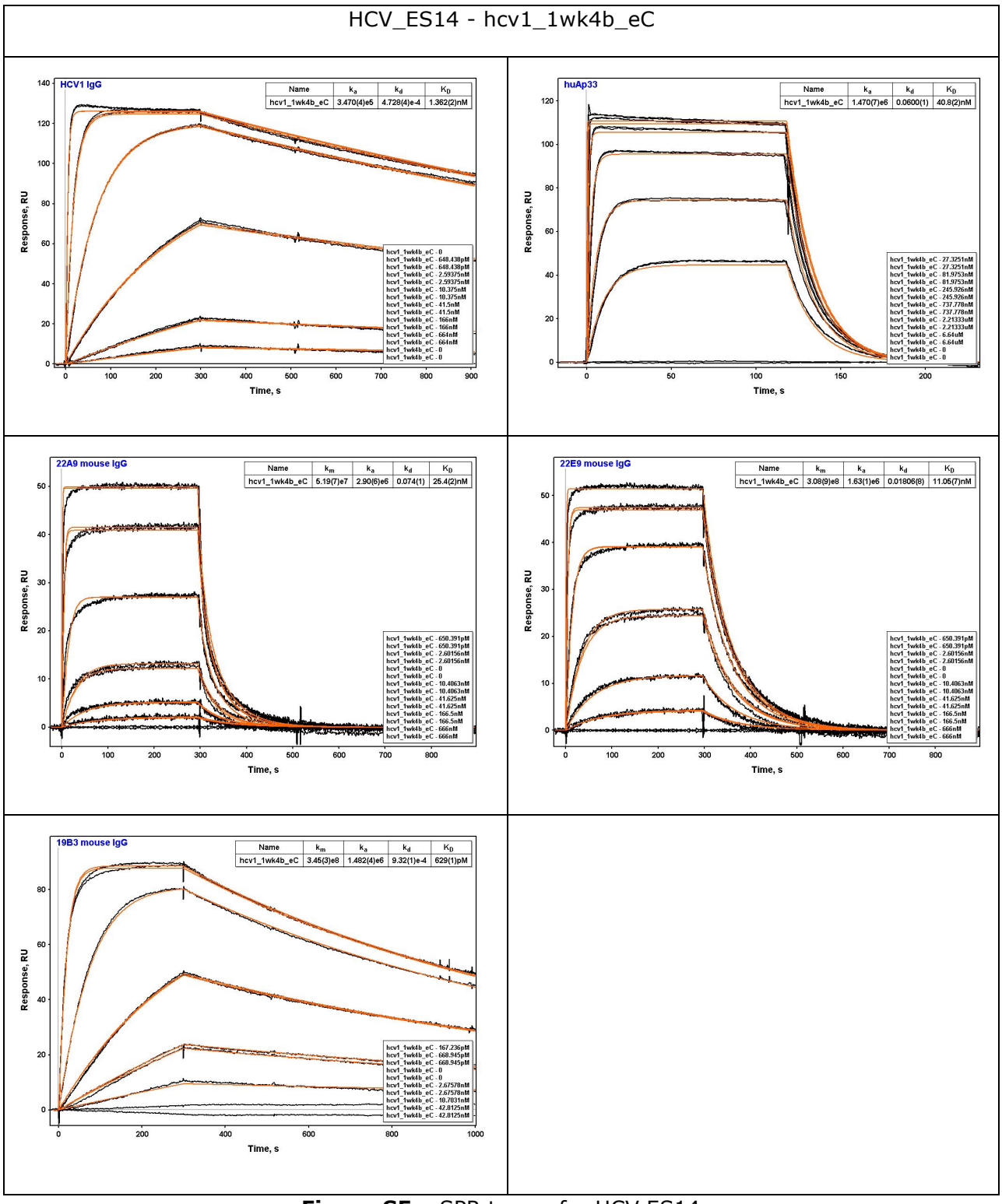


Figure S5 – SPR traces for HCV ES14

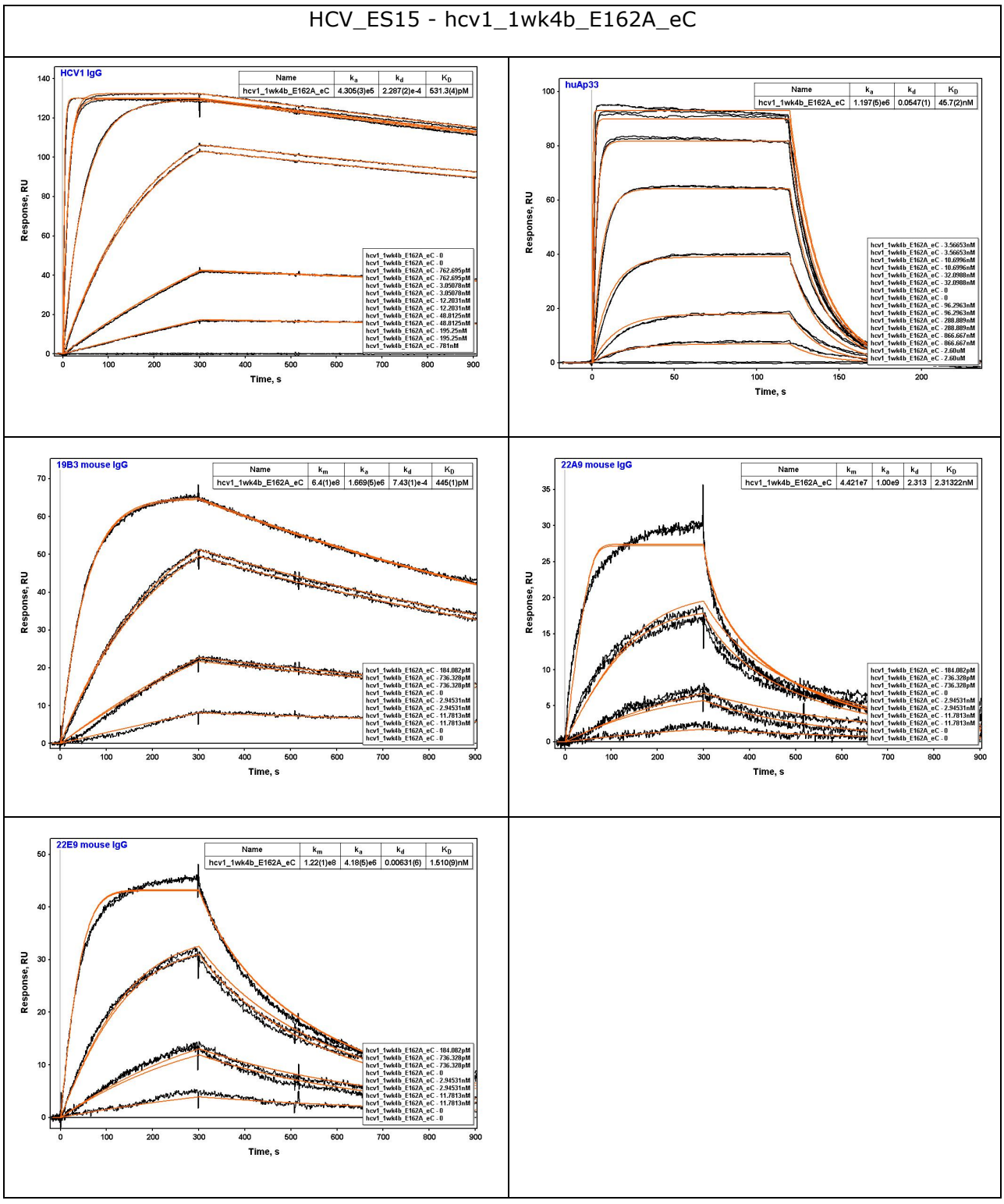


Figure S6 – SPR traces for HCV ES15

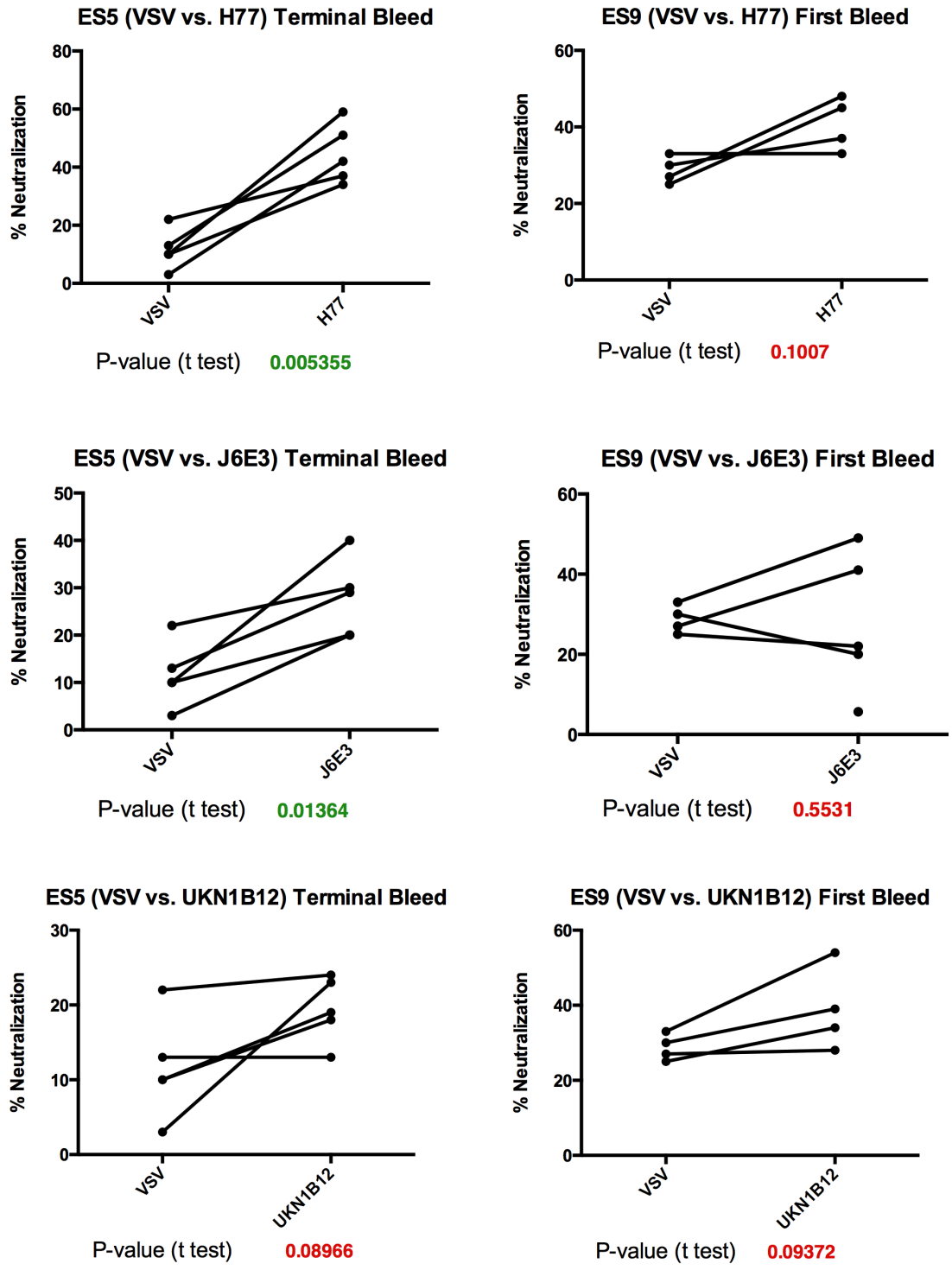


Figure S7 – Paired two-tailed T-test for ES5 and ES9 Immunized Groups. VSV vs. HCV strain at last available timepoint (Terminal Bleed for ES5; First Bleed for ES9). P-value as reported by GraphPad software. P-value less than 0.05 are green, otherwise red.

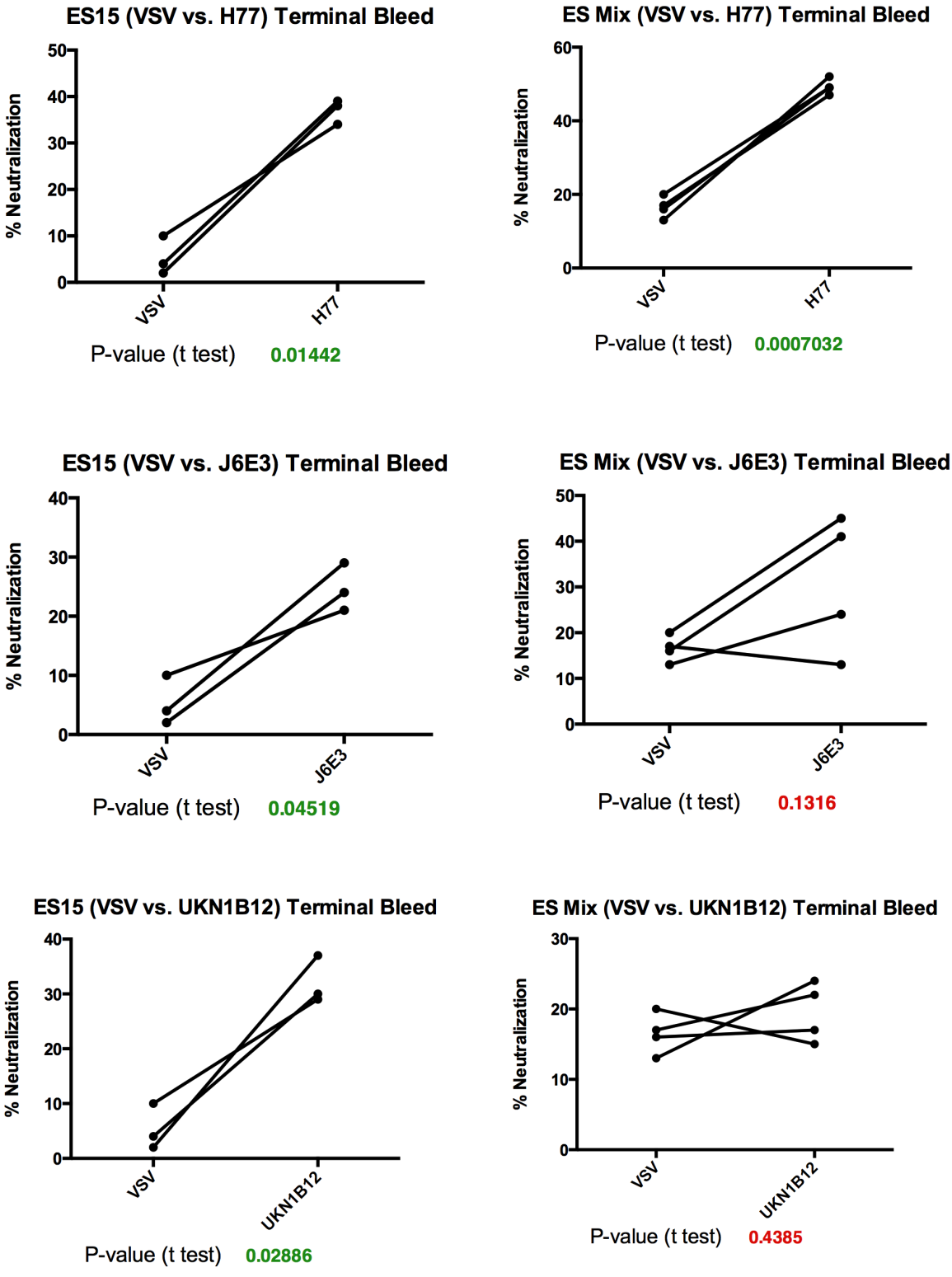


Figure S8 – Paired two-tailed T-test for ES15 and ES Mix Immunized Groups. VSV vs. HCV strain at last available timepoint (Terminal Bleed for ES5 and ES Mix). P-value as reported by GraphPad Prism software. P-value less than 0.05 are green, otherwise red.

CHAPTER 3: PG9 EPITOPE SCAFFOLDING

The HCV case study, presented in Chapter 2, demonstrated that epitope scaffolding is a viable subunit vaccine design strategy, capable of eliciting neutralizing antibodies. Why certain epitope scaffolds fail to elicit neutralizing antibodies remains to be an open question. Accessibility of the epitope in its native environment seems to be a key factor but our sample size of successes is too small to make any strong conclusions. Therefore, the safest strategy, maximizing our chances for success, is to scaffold all known protective epitopes and test their immunogenic properties. In light of this, we chose to continue our scaffolding efforts of protective HIV epitopes.

Similarly to HCV, the surface of the HIV envelope is covered by viral spikes. Each spike is a heterotrimeric assembly of two, glycosylated, proteins gp120 and gp41. It is thought that the high density of glycosylation on gp120 acts as a shield preventing or restricting antibody binding to conserved epitopes (*83*). While self-sugars are generally immunosilent, it is not unexpected that glycans can play an important role in anti-HIV epitopes. In addition to direct interactions in an antibody-epitope interface, the glycans can influence the conformation and steric accessibility of the protein backbone and side-chains.

Recently, there has been a large number of antibodies isolated, which can be described as glycan-dependent (Fig. 3.1) (*84*). Crystal structures of antibody-epitope complexes are not available for all of them, so we cannot definitively state whether the antibodies interact with the glycan directly or if the glycans impose a favorable protein conformation allowing the antibodies to bind.

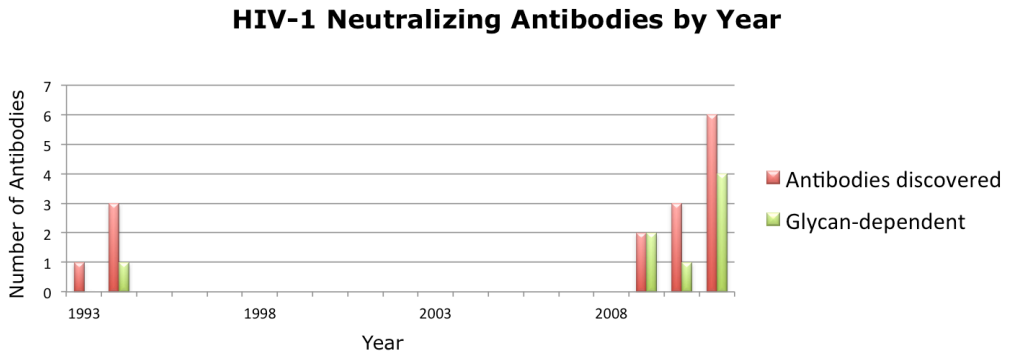


Figure 3.1 – History of HIV antibody discovery: number of broadly neutralizing antibodies discovered per year in red, glycan-dependent antibody count is green. The chart was generated based on a review by Clapham and Lu (*84*).

However, two recently described antibody-epitope complexes reveal a common mode of binding. In the case of PG9 and PGT128, the antibody recognizes a short linear stretch of protein backbone, flanked by two glycans each. Both antibodies were crystallized in complex with engineered constructs designed to present the native epitopes. PG9 (Figure 3.2A) was crystallized in complex with a protein fusion (PDBID: 1FD6) displaying the V1/V2 region of HIV gp120. PGT128 (Figure 3.2B) was crystallized in complex with an engineered outer domain presenting a short V3 loop. In both cases, the core of the protein does not interact with the epitope. We theorized that scaffolding the epitope to stabilize the conformation observed in these crystal structures would be beneficial for reeliciation of such glycan-dependent antibodies. Using these epitopes, we hoped to extend the epitope scaffolding methods to handle glycan-dependent epitopes, which are readily presented on the surface of the wild-type virion and should be the most accessible.

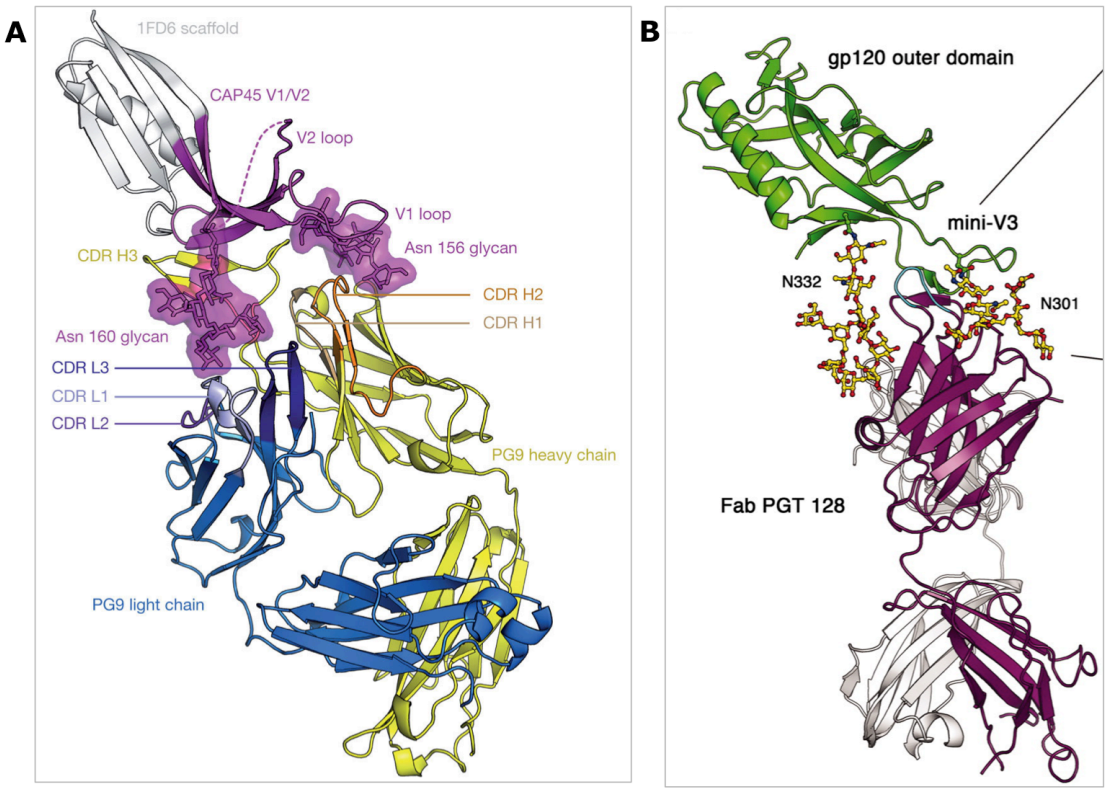


Figure 3.2 – Comparison of PG9 (Panel A) and PGT128 (Panel B) epitopes. Reprinted with permission from Nature Publishing Group (*39*) and AAAS Science Magazine (*53*).

3.1 The PG9 Epitope

The first epitope we attempted to scaffold was that of PG9. Scaffolding efforts for PGT128 were carried out, in parallel, by Ron Jacak, in the Schief Lab.

Transplantation of linear segments has been successfully carried out for several target epitopes. The PG9 epitope is complicated by the presence of glycans, in fact, specific conformations thereof. We theorized that stabilizing the backbone, carrying the glycosylation motif, in the antibody-bound conformation, should guarantee accurate recapitulation of the necessary glycans, in the correct conformation. Given this logic, the problem was simplified to linear epitope scaffolding with high precision.

The best first-generation PG9 scaffold (PDB ID: 1JO8) binds to PG9 with an apparent affinity of 300 nM. The 1FD6 scaffold has an apparent affinity of 3 μM for PG9. If we ignore the glycans, the protein segment of the PG9 epitope is a simple hairpin, which is a common secondary structure motif found in protein structures. We expected to get a large number of prospective scaffolds. The 1FD6 scaffold was crystallized with two different strains of HIV – CAP45 and ZM109.

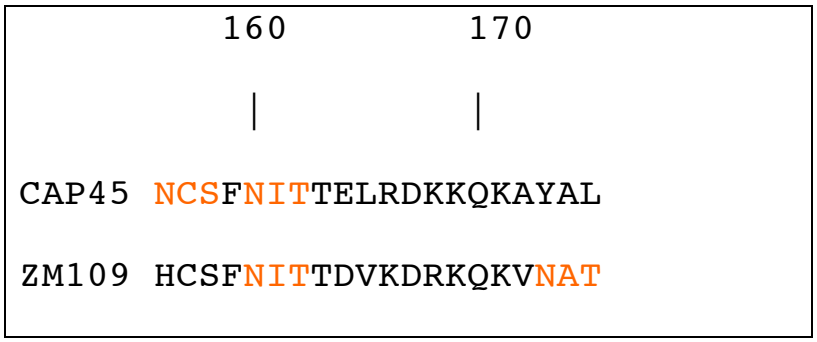


Figure 3.3 – Comparison of HIV strain CAP45 and ZM109, residues 156-175. Putative glycosylation motifs are highlighted in orange

Both CAP45 and ZM109 sequences carry two glycosylation motifs. They are spaced quite differently in sequence but are able to satisfy the same geometric constraints necessary to bind PG9. Glycan at N160 is critical for binding, while, the removal of the second glycan at N156 (CAP45) or N173 (ZM109) significantly reduces affinity for PG9. Therefore, in order to recapitulate both glycans, the second-generation scaffolds should stabilize the full segment from residue 156 to 173.

3.2 PG9 Scaffolding

Using the methods described in the previous chapter, we searched the curated subset of high-resolution structures, from the PDB, for proteins with backbone segments matching

residues 156-173 to identify CAP45-like matches, using the PG9-1F6D-CAP45 complex (PDBID: 3U4E) and residues 160-173 for ZM109-like matches, using the PG9-1F6D-ZM109 complex (PDBID: 3U2S). In the case of ZM109, we did not extend the segment to match the full glycosylation motif, as it is relatively easy to incorporate a serine or a threonine at the design stage to complete the N-x-S/T glycosylation motif.

We identified several prospective matches, which mimic the bound conformation of the PG9 epitope very well. Both CAP45-like and ZM109-like matches were considered. We identified a number of scaffolds, two of which were near-identical matches for the epitope (Table 3.1).

PDB ID:Chain	Protein	Length	Organism
1E4F:T	CELL DIVISION PROTEIN FTSA	378	*T. MARITIMA*
2ZPL:A	REGULATOR OF SIGMA E PROTEASE	97	*E. COLI*

Table 3.1 – Proteins selected as potential scaffolds

We designed a panel of 24 variants, based on the two scaffolds, to present CAP45 and ZM109 sequence (Table 3.2) carrying three types of modifications: 1) point mutants, to resolve any steric clashes with the antibody 2) circular permutations and additional disulfides, to aid with protein stability and 3) domain trims, to reduce the scaffold size and the number of irrelevant epitopes. The FtsA protein (PDBID: 1E4F) was chosen despite its large size. At 378 amino acids, 1E4F is one of the largest scaffolds used in our lab. Previous work with this protein revealed that it can successfully fold even if domains are trimmed off (*85*). Given this information, we were optimistic to design multiple variants, progressively trimming domains, thereby, reducing the size of the scaffold and the number of potentially distracting epitopes. Removing domain 1C, from 1E4F, reduces the scaffold size by 78 residues. Further deletion of domain 1A reduces the scaffold to 162 amino amino acids. Both scaffolds are amenable to circular permutation, utility of which was explored in several

variants. The final set of designed scaffolds is composed of subvariants of four sizes: 1E4F – 378, 311 and 162 residues and 2ZPL – 91 residues (Figure 3.4).

3.3 Expression and biophysical characterization

All of the designs were synthesized by GenScript in pHLsec vector and expressed in 293S cells. After nickel purification, the samples were further purified by size-exclusion chromatography on a Superdex 75 column. 13 out of the 24 scaffolds expressed and were purifiable, specifically ES01-ES06, ES17-ES20, and ES22-24. As described previously in literature, 1E4F was folded even after domain 1C was trimmed. Modifications to 2ZPL were less aggressive and hence most variants based on that scaffold expressed, including the circular permutations of the original protein.

Here, I present the data for PG9 scaffold ES01, as a representative scaffold from the set. This is largely an unmodified protein where epitope positions were transferred onto the native protein scaffold without further backbone perturbation. This scaffold carries the CAP45 epitope sequence and is monomeric and well-folded in solution, as determined by SEC-MALS (Fig. 3.5A), CD wavelength scan (Fig. 3.5B) and thermal melt (Fig. 3.5C). The molecular weight reported by SEC-MALS exceeds the predicted molecular weight by 4 kDa, which can be accounted by the presence of two glycans (approximately 1.8 kDa each) expected for this scaffold. Further, mass spectrometry analysis by Max Crispin at the University of Oxford confirmed the presence of glycans in this sample. Despite this, all of the designed variants failed to bind PG9 with any detectable affinity in an SPR experiment with monomeric gp120 positive control, known to be recognized by PG9.

3.4 Discussion

As expected, we can reliably transplant linear epitopes onto protein scaffolds and obtain well-folded, stable molecules, as illustrated by PG9 ES01. At this intermediate stage, it is too early to draw conclusions, but what is immediately clear is that starting with such a limited set of base scaffolds is a risky strategy for the first round of designs. Removing the glycosylated epitope from its native environment will certainly affect the way it will be glycosylated. Further work is needed to understand the effect of scaffold size and location of the epitope within a given scaffold on glycan modification and processing. A crystal structure of one of the well-behaved scaffolds would be extremely beneficial to answer whether the key asparagine residues in the glycosylation motif were transplanted correctly. Computational methods to model the glycan movements, in context of the scaffolds, are actively being developed. In the meantime, while methods like GlycanRelax (*86*) are still in their infancy, testing a higher numbers of scaffolds should certainly increase the chances of capturing the proper conformation of glycans and protein, in order to see productive binding for PG9, and is the first course of action to be examined.

Failure of first generation scaffolds to bind PG9 highlights the complexity of glycan-dependent epitopes and reiterates the importance of developing appropriate computational methods. Presently, we are able to produce well-folded and specifically glycosylated molecules. However, it is unclear whether the glycans exist in the proper conformation to bind their target antibodies. We will attempt to crystallize available first generation scaffolds in addition to testing a larger set of base scaffolds, in hopes of identifying a scaffold compatible with the necessary glycan conformations.

Base Protein	Sequence scaffolded	Scaffold Name	Additional optimizations
1e4f	CAP45	PG9_ES01	
1e4f	CAP45	PG9_ES02	Y284G
1e4f	CAP45	PG9_ES03	graft
1e4f	CAP45	PG9_ES04	graft, Y284G
1e4f	CAP45	PG9_ES05	domain 1C trim
1e4f	CAP45	PG9_ES06	domain 1C trim, Y284G
1e4f	CAP45	PG9_ES07	graft, domain 1C trim
1e4f	CAP45	PG9_ES08	graft, domain 1C trim, Y211G
1e4f	CAP45	PG9_ES09	domain 1A trim
1e4f	CAP45	PG9_ES10	domain 1A trim, Y93G
1e4f	CAP45	PG9_ES11	graft, domain 1A trim
1e4f	CAP45	PG9_ES12	graft, domain 1A trim, Y93G
1e4f	CAP45	PG9_ES13	domain 1A trim, circ. permutant
1e4f	CAP45	PG9_ES14	domain 1A trim, circ. Permutant, Y93G
1e4f	CAP45	PG9_ES15	graft,domain 1A trim, circ. permutant
1e4f	CAP45	PG9_ES16	graft,domain 1A trim, circ. Permutant, Y93G
2zpl	ZM109	PG9_ES17	
2zpl	ZM109	PG9_ES18	P57T
2zpl	ZM109	PG9_ES19	extra glycan
2zpl	ZM109	PG9_ES20	graft, V67G
2zpl	ZM109	PG9_ES21	graft, extra glycan
2zpl	ZM109	PG9_ES22	graft, circ. permutant
2zpl	ZM109	PG9_ES23	circ. permutant, V85G
2zpl	ZM109	PG9_ES24	graft, circ. permutant, extra glycan

Table 3.2 – PG9 Scaffold Variants: original protein, the sequence variants (CAP45 or Z109) displayed, short name for the scaffold and changes from the standard superposition protocol.

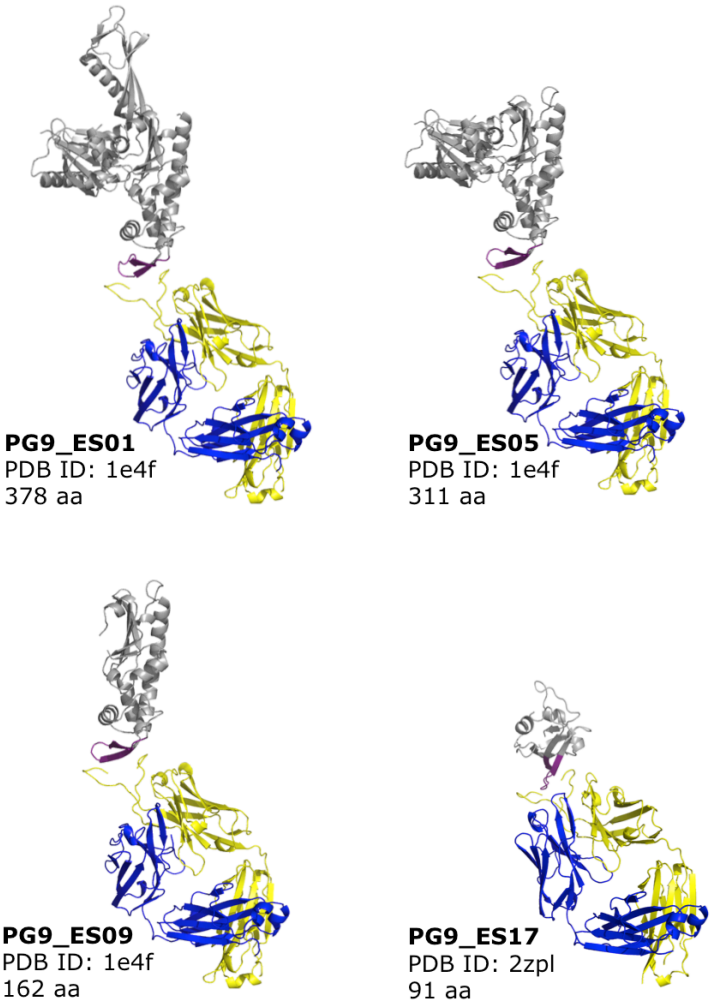


Figure 3.4 – Overview of base scaffolds and their trimmed versions. PG9 antibody light chain – blue, heavy chain – yellow. Scaffold – gray. PG9 epitope – purple.

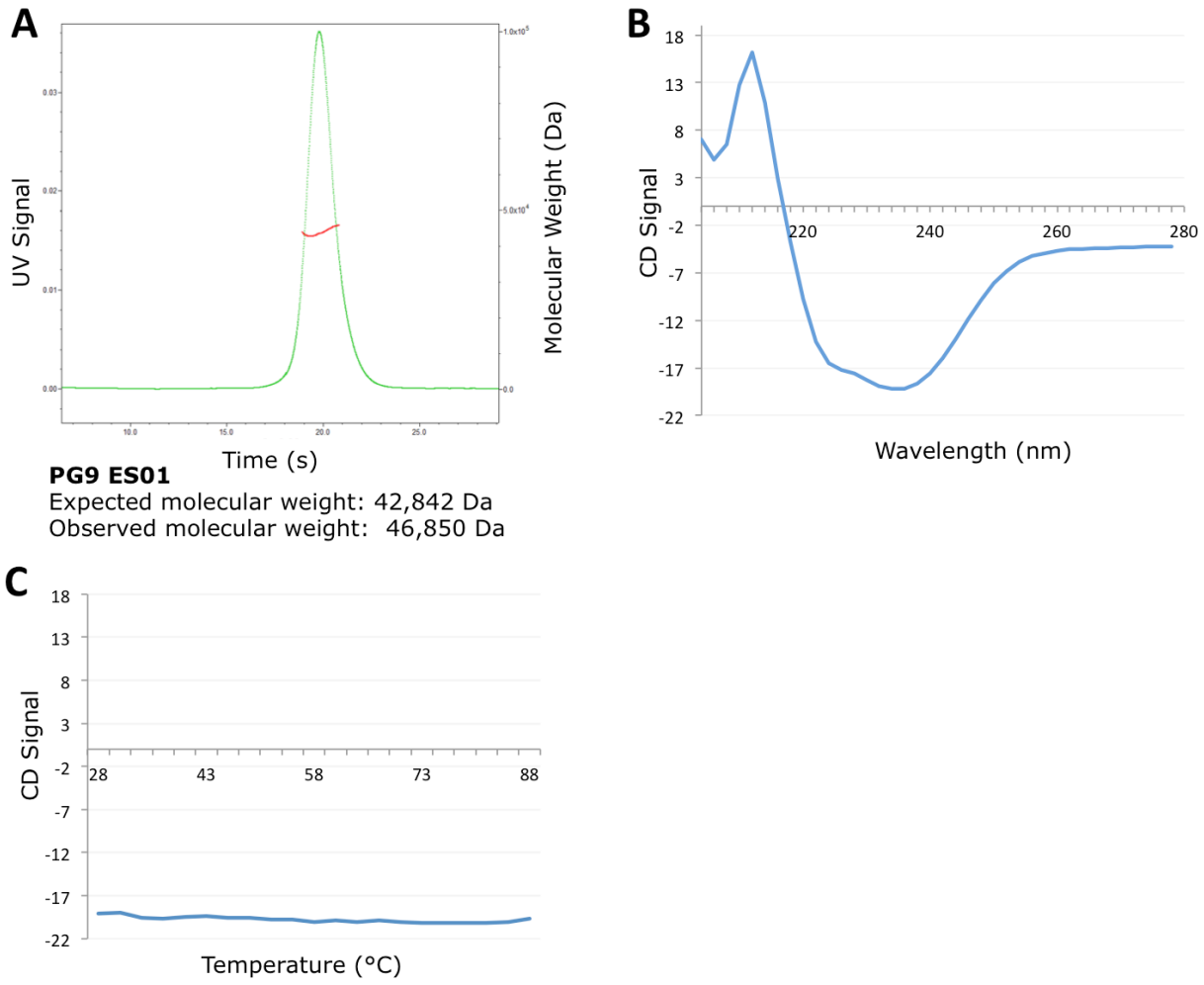


Figure 3.5 – Representative biophysical profile for PG9 ES01 scaffold (based on 1E4F, scaffolding CAP45 sequence). A) SEC-MALS analysis for ES01 – scaffold is monomeric and elutes as a single peak (UV trace – green) and predicted molecular weight (red) is in agreement with predicted values (+/- 2 glycans at 1.8 kDa each) B) CD wavelength scan (blue) has a characteristic alpha-helical, folded protein shape at room temperature C) Thermal stability of PG9 ES01 monitored at 222 nm.

CHAPTER 4: MULTIVALENT PRESENTATION PLATFORMS

Chapter 2, the HCV1 epitope scaffolds have demonstrated that we can make stable immunogens capable of eliciting neutralizing response for HCV. Chapter 3 highlighted the difficulty of scaffolding glycan-dependent epitopes in the context of HIV. Conceptually, the native protein is the best possible scaffold for any given epitope. In the case of HIV, the gp120 envelope glycoprotein ideally presents the CD4 binding site (CD4bs), which is targeted by several bNAbs, as well as all of the known glycan-dependent bNAb epitopes. Immunization with gp120 fails to reelicit such antibodies due to multiple immunodominant off-target epitopes. The outer domain (OD) of gp120 has been extensively reengineered to remove these epitopes, resulting in a minimal immunogen (eOD) accurately mimicking the CD4bs and some of the glycan-dependent epitopes (Huang et al., in preparation). The HCV1 scaffolds, for HCV, and the eOD construct, for HIV, are both promising immunogens. To further improve the immunogenicity of this molecule, I developed a general multimerization platform using the eOD immunogen as a proxy for future epitope scaffolds. The platform developed was used to display a germline-targeted version of eOD in a polyvalent fashion, which has been a collaboration of Joseph Jardine, myself and Jean-Philippe Julien published in Science (*32*).

4.1 Introduction

Protection against disease by nearly all licensed vaccines is associated with induction of antibodies (*1*). Viruses with high antigenic diversity, such as HIV, influenza virus, and hepatitis C virus, pose major challenges for vaccine development (*87*). Most exposed surfaces on the Envelope glycoproteins (Env) of these viruses are hypervariable or shielded by glycans (*88*), and traditional vaccine approaches tend to induce neutralizing antibodies against only a small subset of viral strains (*14*, *17*, *89*) However, discoveries of bNAbs against each of these viruses have identified conserved epitopes as leads for vaccine design

(87), and structural analysis has provided atomic definition for many of these epitopes (5, 90). Structure-based approaches are, therefore, needed to reverse-engineer vaccines capable of inducing bNAbs against these conserved epitopes (35).

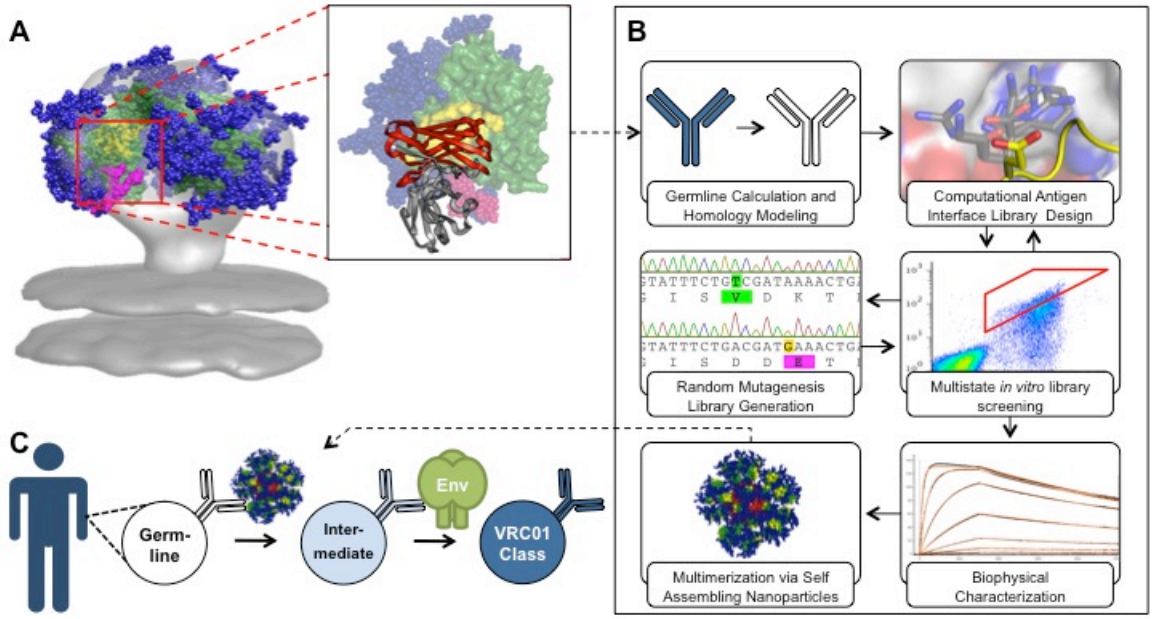


Fig. 4.1. Development of a germline (GL)-targeted HIV immunogen.

(A) VRC01-class bNAbs bind to gp120 primarily through paratope residues encoded by VH 1-2*02. gp120 is colored green with the CD4 binding site highlighted in yellow. Glycans are represented as blue spheres with the critical N276 highlighted in magenta. VRC01 is shown as secondary structure rendering and colored gray, with the VH1-2*02 region highlighted in red. (B) Steps in the engineering of a modified gp120-based nanoparticle capable of activating GL VRC01-class B cells. (C) This nanoparticle can be used in an HIV-1 vaccine GL-prime-boost strategy to bridge this initial recognition gap and initiate clonal expansion and start somatic hypermutation of VRC01-class bNAbs precursors.

High potency VRC01-class bNAbs against the HIV gp120 CD4 binding site (CD4bs) have been isolated from several individuals infected with different strains of HIV-1 (29-31). VRC01-class bNAbs all derive from the human VH1-2*02 variable heavy gene, but differ significantly in amino acid sequence and complementarity-determining region H3 (CDRH3)

length and use a few different variable light chain genes (figs. S1-2). Structural studies have revealed that VRC01-class bNAbs employ a common mode of gp120 binding in which the VH1-2 framework mimics CD4 and provides additional electrostatic and hydrophobic contacts (Fig. 4.1A) (*30*, *91-93*). A short CDRL3 loop is also required for interaction with gp120 V5 and Loop D, and a CDRL1 deletion in many VRC01-class bNAbs avoids clashes with a glycan linked to Asn276 (N276) on loop D.

Vaccine design to induce VRC01-class bNAbs is attractive because VH1-2 genes are estimated to be present in ~2% of the human Ab repertoire (*94*) and, even considering restrictions on light chain usage, suitable precursors should be present in the naïve B cell repertoire of most individuals. However, predicted germline (GL) precursors for VRC01-class bNAbs exhibit no detectable affinity for wild-type Env (*31*, *91*) (Table 4.1 and table S1), a potential explanation for the rarity of VRC01-class bNAbs in HIV-1 infection (*91*). More importantly, wild-type Env constructs lacking GL affinity are poor vaccine candidates to prime VRC01-class responses, as they are unlikely to reliably stimulate GL precursors to initiate antibody maturation.

	Antigen	Antibody															
		VRC01		12A12		3BNC60		NIH45-46		PGV04		PGV19		PGV20		VRC-CH31	
		GL	Mat	GL	Mat	GL	Mat	GL	Mat	GL	Mat	GL	Mat	GL	Mat	GL	Mat
WT CD4 Binding Site	Core.HXB2	>10^5	5	>10^5	6	>10^5	36	>10^5	35	>10^5	48	>10^5	20	>10^5	19	>10^5	47
	eOD-Base N276D	>10^5	5	>10^5	380	>10^5	4,100	>10^5	14	>10^5	110	>10^5	3,100	17,000	16	>10^5	30,000
Modified CD4 Binding Site	Core.BaL-GT1	1,800	0.5	3,200	1	16,400	4	>10^5	0.6	>10^5	170	25	14	5,500	4	35,000	1,800
	eOD-GT1	44,000	1	>10^5	2,300	>10^5	83	>10^5	3	>10^5	4	7,800	1,000	1,100	10	>10^5	>10^5
	eOD-GT6	44	2	2,000	400	14,000	200	410	4	52,000	10	19	88	3	6	28,000	29,000

Table 4.1. Binding of GL and mature (Mat) Abs to gp120 and eOD variants. Values represent K_Ds in nM measured by SPR. Detectable binding to GL antibodies is highlighted in red.

4.2 Immunogen Design Strategy

To address the problem described above, we modified the CD4bs on a minimal, "engineered outer domain (eOD)" (*53*) to produce a "germline-targeting" vaccine prime (Fig. 4.1) with two important binding properties: (i) moderate affinity for multiple predicted VH1-2*02 GL-Abs to enhance the ability to activate VH1-2 GL B cells with appropriate light chains; (ii) high affinity for VRC01-class bNAbs to provide an affinity gradient to guide early somatic mutation toward the mature bNAbs. Furthermore, we developed self-assembling nanoparticles presenting 60 copies of the germline-targeting eOD, to enhance B cell activation and to improve trafficking to lymph nodes (Fig. 4.1).

4.3 Engineering and Biophysical Analysis of Germline-Targeting Antigens

Modifications to the VRC01 epitope included removing clashes and building new contacts between the CD4bs and the GL-Abs, as well as rigidifying the CD4bs in a conformation that is favorable for binding. Initially, we constructed a homology model of GL-VRC01 bound to gp120 and identified a likely clash between CDRL1 and the N276 glycan. Therefore, we evaluated the GL-VRC01 binding of an eOD (eOD-Base) that lacks glycans at 276 and on the nearby V5 loop owing to N276D and N463D mutations. The eOD-Base barely interacted with GL-VRC01 (K_D ~ 1 mM) and had low affinity for only 2 of 8 other GL VH1-2*02 Abs tested (Table 4.1). We then used Rosetta computational protein interface design (*95*) to identify other mutations in and around the CD4bs that were predicted to improve GL-VRC01 binding, and created directed libraries that included all possible combinations of the computationally identified mutations. These libraries were screened for gp120 core and eOD variants that showed GL-VRC01 binding using yeast cell surface display (*96*). This strategy generated germline-targeting (GT) variants of gp120 core (CoreBal-GT1) and eOD (eOD-GT1) that bound GL-VRC01 with K_Ds of 1.8 μM and 44 μM, respectively (Table 4.1 and table S2). We focused further development on the smaller eOD because it lacks potentially distracting epitopes on gp120 core. A second round of computational design and directed-library screening produced eOD-GT2 with a three-fold improvement in K_D for GL-VRC01 (table S2). Subsequent screening of mixed computational/random mutagenesis libraries led to larger improvements and resulted in eOD-GT3 and eOD-GT4, which had K_Ds for GL-VRC01 of 220 and 34 nM, respectively (table S2). Interestingly, screening for GL-VRC01 binding also improved binding to other GL VH1-2 Abs, as eOD-GT4 bound to GL-NIH45-46 and GL-PGV19 with K_Ds of 1.0 μM and 28 nM, respectively (table S2). To achieve these improvements, eOD-GT4 had accumulated 17 mutations relative to eOD-Base.

To retain as native a CD4bs as possible (by reducing the number of mutations) while also maintaining or improving binding to GL VH1-2 Abs and mature bNAbs, we employed multi-target optimization. Here, libraries with either the wild-type HIV-1 strain HxB2 residue or the mutation in eOD-GT4 were sorted in parallel against multiple Abs, and mutations were retained only if they were positively selected by at least one GL Ab and not negatively selected by other Abs used during optimization. eOD-GT6 was generated by sorting against GL Abs for VRC01, NIH45-46, PGV19, PGV04 and VRC-CH31, as well as mature VRC01 and PGV04. eOD-GT6 had only 8 mutations relative to eOD-Base (10 mutations relative to HxB2 gp120) and retained excellent binding to diverse GL VH1-2*02 Abs, with a K_D of 44 nM for GL-VRC01 and K_Ds < 500 nM for 5 of 9 GL Abs tested (Table 4.1, table S2, fig. S3). eOD-GT6 also had high affinity for several mature bNAbs, with K_Ds of 2, 4, and 88 nM for VRC01, NIH45-46, and PGV19, respectively, and maintained the desired affinity gradient for 6 of 8 Abs tested (Table 4.1). Further, eOD-GT6 bound with high affinity to GL VRC01-class Abs derived from VH1-2*03 and *04 alleles (table S3, fig. S4). eOD-GT6 had no detectable affinity for VH1-2*01, probably due to the absence of Trp^{H50}; however, recent data from sequencing of 1092 human genomes (*97*) shows that the W50R mutation in VH1-2*01 occurs at a frequency of 0.21, indicating that only ~4% of the population are VH1-2*01 homozygotes not amenable to eOD-GT6 priming.

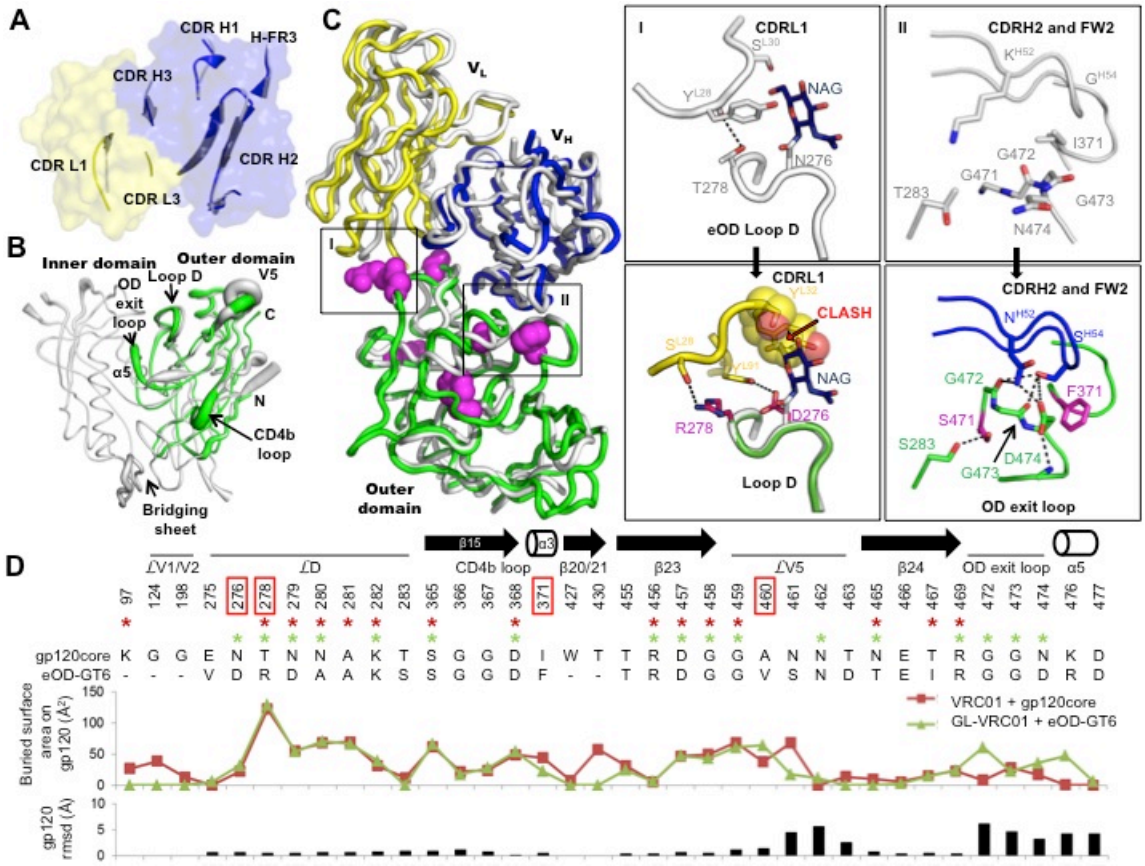


Fig. 4.2. Structural analysis of GL-VRC01 and eOD-GT6.

(A) The structure of the unliganded inferred GL-VRC01 antibody (heavy and light chains colored blue and yellow, respectively) is similar to the structure of gp120-bound VRC01 (white) within the gp120-contacting positions (shown in orange on VRC01, defined by the structure of VRC01+gp120 in PDBID:3NGB). Structures are rendered according to B-values, with thin and thick lines representing areas possessing low and high flexibility, respectively. (B) Comparison between the crystal structures of unliganded eOD-GT6 (green) and unliganded gp120 core from HIV-1 strain 93TH057 (PDBID: 3TGT, white). Structures rendered as in (A). (C) Comparison between the crystal structures of GL-VRC01+eOD-GT6 and VRC01+gp120 core (PDBID: 3NGB), in which only the outer domain of gp120 is shown. Structures rendered as in (A) and (B), except gp120-contacting positions on VRC01 are white. The mutated residues in eOD-GT6 that enable binding of GL-VRC01-class Abs are shown as space-fill magenta spheres. The angle of approach of GL-VRC01 and VRC01 to the CD4bs is nearly identical. Key regions where interactions are different between VRC01 on gp120 (upper panels) and GL-VRC01 on eOD-GT6 (lower panels) are shown in insets. eOD-GT6 confers germline reactivity by removing a potential clash with the N276 glycan, as well as by creating additional contacts with loop D (left panels), the OD exit loop (right panels) and V5 (fig. S12). (D) gp120 residues involved in the VRC01+gp120 and GL-VRC01+eOD-GT6 interfaces are compared in sequence, H-bond (stars), surface buried area and RMSD. Interfaces were calculated using PDBePISA (*98*) and Cα rmsd using Chimera (*99*).

4.4 Crystallographic Analysis

To understand the molecular interaction of eOD-GT6 with GL-VRC01, we solved three crystal structures: unliganded GL-VRC01, unliganded eOD-GT6, and the complex of GL-VRC01 bound to eOD-GT6, to resolutions of 2.1 Å, 2.5 Å, and 2.4 Å, respectively (Fig. 4.2 and table S4). The unliganded GL-VRC01 structure revealed that the gp120 contacting loops closely resemble those of VRC01 despite extensive affinity maturation of the latter (Fig. 4.2A, and figs. S5-7). Unliganded eOD-GT6 showed a similar structure to the outer domains of unliganded and VRC01-bound gp120 core (1.2 Å Cα root-mean-square-deviation (RMSD) in both instances) (Figs. 2B and fig. S8), suggesting good mimicry of the CD4bs. The structure of unliganded eOD-GT6 was also similar to eOD-GT6 bound to GL-VRC01 Fab (0.9 Å RMSD), with the largest differences occurring in the flexible loops and in the eOD exit loop (fig. S9). In addition, the structure of the eOD-GT6+GL-VRC01 complex indicated that the VH1-2-encoded domain of GL-VRC01 approaches eOD-GT6 at an angle nearly identical to that of VRC01 with gp120 (Fig. 4.2C) (4.2° angular difference when the complexes are superposed on the CD4 binding loop). Overall, the buried surface area (BSA) of GL-VRC01 (1076 $Å^2$) on eOD-GT6 (1102 $Å^2$) is nearly identical to VRC01 (1152 $Å^2$) on gp120 core (1120 $Å^2$), further demonstrating the high degree of similarity between the two structures (tables S5-6). Key hydrogen-bonding networks are preserved in the GL-VRC01+eOD-GT6 interaction, particularly in the CD4 binding loop (fig. S10). On the other hand, important differences in hydrogen bonds, BSA, and Cα positions are observed for interactions in loop D, V5 and the OD exit loop, which contribute to GL-VRC01 reactivity to eOD-GT6 (Fig. 4.2D and fig. S11).

4.5 Mutation Analysis

To understand the affinity contributions of individual mutations, we measured GL-VRC01 binding affinities for point reversions of each mutation (Table 4.2). Six mutations on eOD-GT6 conferred improved affinity for GL-Abs relative to the starting construct (eOD-Base) that lacked glycans at 276 and 463. The eOD-GT6+GL-VRC01 complex structure revealed that these mutations are either directly involved in the binding interface (T278R, I371F and N460V) or stabilize loops involved in the interface (L260F, K357R and G471S) (Fig. 4.2C). The two mutations with the largest effect on GL-VRC01 binding were G471S and I371F; reversion at these positions reduced GL-VRC01 affinity by factors of 39 and 10, respectively (Table 4.2). Ser471, together with Phe371 and Phe260, appears to play a role in altering the conformation of the OD exit loop to allow the GL-VRC01 CDRH2 to make H-bonds with three additional gp120 residues (G472, G473 and D474) and bury an additional 120 $Å^2$ on gp120, resulting in improved binding (Fig. 4.2C-D, right inset panel, tables S5-6, fig. S9). Also, the N460V mutation located in V5 improves packing with the antibody and appears to contribute to an altered V5 conformation and pattern of V5 H-bonding with VRC01, as compared to Clade A/E 93TH057 gp120 recognition of VRC01 (Fig. 3.2D and fig. S12). Reversion of the N460V mutation reduced GL-VRC01 binding by a factor of 2.5 (Table 4.2).

	eOD-GT6	Single Reversions					
Reversion from eOD-GT6	-	F260L	R278T	R357K	F371I	V460N	S471G
Frequency of WT AA	-	L (97.3%)	T (78.4%)	K (68.9%)	I (85.8%)	N (36.2%)	G (78,7%)
Frequency of Mutation	-	F (0.0%)	R (0.1%)	R (0.5%)	F (0.0%)	V (3.4%)	S (0.5%)
Method of Discovery	-	epPCR	Rosetta	epPCR	Rosetta	Rosetta	epPCR
GL-VRC01 Affinity (nM)	44	312.7±0.4	156.2±0.2	64.2907±0.0009	454±1	119±2	1,670±10
Change from eOD-GT6	-	7.0	3.6	1.5	10.3	2.7	39

Table 4.2. Binding of GL VRC01 to eOD-GT6 point reversions. Values represent K_Ds in nM measured by SPR. Amino acid frequencies determined from 2867 HIV-1 sequences from http://www.hiv.lanl.gov/.

Removal of key glycosylation sites was necessary for GL affinity. Reintroduction of the N276 glycosylation site in eOD-GT6 (by a double mutation, D276N/R278T) reduced binding by a factor of 140, and the remaining binding was likely due to a small fraction of the eOD-GT6-D276N/R278T that underutilized the N276 glycosylation position (fig. S13). Reversion of R278T alone reduced affinity by a factor of only 3.6 (Table 4.2). Thus, removal of the 276 glycan appears to release a block on GL-VRC01 binding, but does not confer appreciable eOD affinity; further interface modification was required to achieve high affinity. Indeed, the eOD-GT6+GL-VRC01 complex structure revealed that, in addition to removing a clash between the N276 glycan and CDRL1 (Fig. 4.2C, left inset panel), eOD-GT6 D276 and R278 make two additional H-bonds with GL-VRC01. eOD-GT6 also lacks glycans at positions 386 (β12) and 463 (V5). Restoration of these glycosylation sites reduced affinity for GL-VRC01 by a factor of 3 (table S7).

4.6 eOD-GT6 Nanoparticle Generation

To enable eOD-GT6 to activate GL B cells via cross-linking of B cell receptors, and to develop a multivalent platform for eOD-GT6 that mimics the size, shape, multivalency and symmetric surface geometry of many viruses for improved immunogenicity (*61*), we sought to fuse eOD-GT6 to a self-assembling virus-like nanoparticle. From a search of large homomeric particles in the Protein Data Bank (PDB), we prioritized 60mers of Lumazine Synthase from the hyperthermophile *Aquifex Aeolicus* for experimental testing due to their thermal stability and because modeling suggested that, with a suitable linker length, 60 copies of glycosylated eOD-GT6 could be sterically accommodated in an orientation that would expose the VRC01 epitope (Fig. 4.3A). Although expression of the wild-type particle had been reported in *E. coli* (*68*), we found that such nanoparticles presenting glycosylated eOD-GT6 could be secreted from mammalian (293) cells and purified by lectin chromatography with good yield (~10 mg/L) and structural homogeneity (Fig 4.3B and figs S14-15).

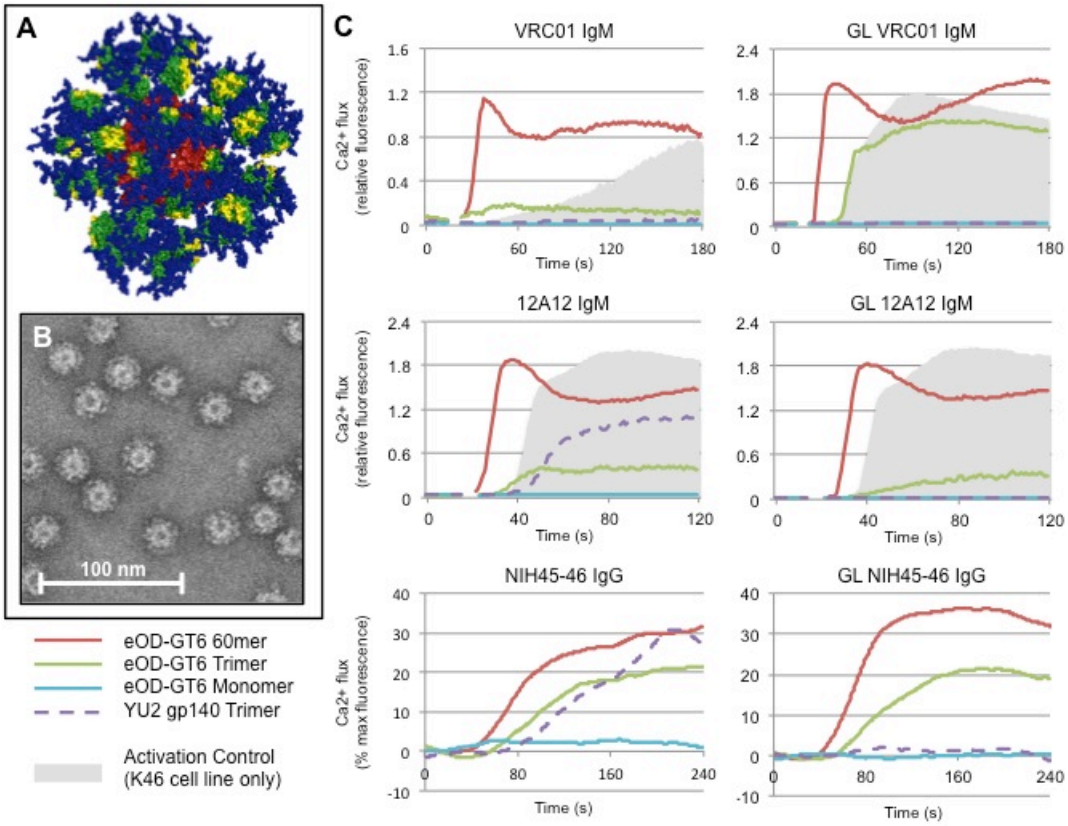


Fig. 4.3. A 60-mer eOD-GT6 nanoparticle activates GL and mature VRC01-class B cells. (A) Model representation of the 60-mer eOD-GT6 nanoparticle. eOD-GT6 is colored in green, with residues that interact with VRC01 colored in yellow. Glycans are shown as blue spheres and the self-assembling 60mer Lumazine Synthase to which eOD-GT6 is fused is colored red. (B) Raw negative stain electron microscopy image of the 60-mer eOD-GT6 nanoparticle. (C) Calcium flux experiments show that the 60-mer eOD-GT6 nanoparticle activates B cell lines engineered to express either GL or mature VRC01 IgM, 12A12 IgM or NIH45-46 IgG, while a recombinant soluble gp140 trimer activates the B cells expressing mature, but not GL VRC01-class Abs. Data for each antibody are representative of at least two separate experiments.

4.7 In Vitro B Cell Activation

The ability of eOD-GT6 nanoparticles to activate B cells expressing GL and mature VRC01 (IgM) (*100*), 12A12 (IgM and IgG) and NIH45-46 (IgG) (*101*), was tested in Ca^{2+}-dependent activation assays. The 60-mers potently activated both GL and mature B cells with 1 μM outer domain (16 nM particle) and modestly activated all three cell lines at 1000-fold lower concentrations (Fig. 4.3C and fig. S16). In contrast, monomeric eOD-GT6 was non-stimulatory, probably due to an inability to cross-link B cell receptors (*100*). Trimeric eOD-GT6 activated both GL and mature B cells, but less potently and rapidly than the 60mers, and a soluble gp140 trimer from HIV-1 strain YU2 (*102*) showed no activation of GL B cells, but did activate the mature counterparts (Fig. 4.3C). Both IgM and IgG B cell lines were generated for GL 12A12 and we observed no significant differences in activation magnitude or kinetics between the two antibody isotypes (fig. S16).

4.8 Animal Models for Human VH1-2 Germline-Targeting

We then assessed whether eOD-GT6 might interact with related GL-Abs in animal models. Analysis of VH genes from rabbit (fig. S17) (*103*), mouse (figs. S18-19) (fig. S19) and macaque (fig. S20) revealed that none of these commonly used model organisms have a known VH gene containing all of the critical residues for GL binding (*93*). To measure binding experimentally, chimeric GL-Abs were produced in which the human VH1-2*02 gene from GL-VRC01 was replaced with GL VH genes from mice or macaques containing the essential Arg^{H71} and as many other critical residues as possible. Chimeric GL-Abs with mouse VH genes had no detectable binding to eOD-GT6. Chimeric Abs derived from two of the three rhesus VH genes bound weakly to eOD-GT6, with K_Ds of ~30 μM and ~40 μM (table S8). The rhesus chimeric GL-Ab most similar to human GL-VRC01 contained only 10

mutations in the VH gene (95.5% identity over the antibody Fv region) but showed no detectable binding to eOD-GT6. Annotation of the rhesus macaque antibody repertoire and analysis of gene usage frequencies will be useful to construct *bona fide* macaque GL VH1-2 Abs. These analyses illustrate the potential difficulty for using animal models to produce VRC01-class bNAbs, and suggest that immunization of humans or mice engineered to produce human Abs may be essential for testing and iteratively optimizing such immunogens.

The events that led to GL VH1-2*02 B cell activation in the HIV-infected individuals from which VRC01-class bNAbs were isolated remain unclear. Our finding that a small number of rare or previously undocumented Env mutations confers high affinity GL binding suggests that Env variants might have acquired one or more such mutations stochastically during infection and thereby gained the ability to prime GL VRC01-class B cells. Vaccines to induce VRC01-class responses will need to activate such B cells dependably and drive appropriate somatic mutation to produce high affinity bNAbs (*104*). We propose the eOD-GT6 nanoparticle as a promising candidate for a vaccine prime based on its ability to bind diverse VH1-2*02 GL Abs, activate VRC01, 12A12 and NIH45-46 B cell lines *in vitro*, and provide an affinity gradient for early somatic mutation. We further propose that ultimate elicitation of mature VRC01-class bNAbs will require, at minimum, boosting with different immunogens that present a less engineered, more native CD4bs including the glycans around the CD4bs.

4.9 Acknowledgments:

We thank J. Mascola and X. Wu for providing sequences of mature PGV19 and PGV20, R. Wyatt for providing the YU2 gp140 trimer, C. Sundling and G. Karlsson Hedestam for the sequence of Rhe1 VH gene, A. Bradley, R.L. Stanfield, and D.C. Diwanji for technical

assistance and insightful discussions, and M. Azoitei, R. Jacak, and D. Kulp for comments on the manuscript. The data presented in this manuscript are tabulated in the main paper and the supplementary materials. Coordinates and structure factors for GL-VRC01 Fab, eOD-GT6 and GL-VRC01+eODGT6 structures will be deposited with the Protein Data Bank and released upon publication. The International AIDS Vaccine Initiative (IAVI) has filed a patent relating to immunogens in this manuscript: PCT Application No. PCT/US12/60062, titled "Engineered Outer Domain (EOD) Of HIV GP120 And Mutants Thereof", with inventors WRS, JJ, SM, PH. Materials and information will be provided under MTA. This work was supported by IAVI Neutralizing Antibody Center, CHAVI-ID (UM1 AI100663), NIH grants AI84817 (I.A.W.), AI081625 (L.S.) and AI33292 (D.R.B.), NIH NRSA Training Grant fellowship T32CA080416 (J.J.), Canadian Institutes of Health Research fellowship (J.-P.J.), CNIHR grant P30 AI027767-24 (T.O.), NIH Interdisciplinary Training Program in Immunology 5T32AI007606-10 (D.S.) and the Ragon Institute. Portions of this research were carried out at the Stanford Synchrotron Radiation Lightsource (SSRL), a Directorate of the SLAC National Accelerator Laboratory and an Office of Science User Facility operated for the U.S. Department of Energy (DOE) Office of Science by Stanford University. The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research; NIH's National Center for Research Resources, Biomedical Technology Program (P41RR001209); and the National Institute of General Medical Sciences (NIGMS). Use of the Advanced Photon Source was supported by the DOE, Basic Energy Sciences, Office of Science, under contract no. DE-AC02-06CH11357. GM/CA CAT has been funded in whole or in part with federal funds from NCI (grant Y1-CO-1020) and NIGMS (grant Y1-GM-1104). This is manuscript 23050 from The Scripps Research Institute.

4.10 Supplementary Materials:

Germline Antibody Analysis

Ab GL Predictions. The naïve heavy and light chain precursors of VRC01-class bNAbs 12A12 (*11*), 3BNC60 (*11*), NIH45-46 (*11*), PGV04 (*12*), PGV19, PGV20, VRC-CHA31 (*12*), VRC01 (*10*), (figs. S1-2) were calculated from the mature Ab sequences using JOINSOLVER® (*31*) to select probable variable (V), diversity (D) and joining (J) genes. In nearly all cases, the genes of highest probability predicted by JOINSOLVER® were selected but, in some cases (such as J gene selection), we chose the second most likely gene because it contained elements that we believed made the selection more biologically relevant (figs. S20-21). For most antibodies (Abs), mature CDR3 loops were used, as it was not possible to unambiguously determine the original identity of the junction region. For some of the Abs (VRC01, NIH45-46, VRC-CH31 and PGV04), several variations of the CDRH3 or light chain (LC) pairings were constructed.

Homology Modeling. Initially, homology modeling of VRC01 and PGV04 germline (GL) Abs based on the mature Ab structures was used to generate starting structures for design. The coordinates of the mature VRC01 and PGV04 were extracted from the co-crystal structures bound to gp120 (PDBID 3NGB and 3SE9, respectively). The two amino-acid deletion in CDRL1 was modeled transferring the CDRL1 from a minimally mutated LC from the same VL gene (PDBID: 1GC1 and PDBID: 3F12) using RosettaRemodel (*32*). After correcting the length change, the GL sequence was threaded onto the coordinates of the mature Ab using RosettaFixBB (*33*) design, and the resulting GL model was relaxed using RosettaRelax (*33*). In these calculations, the constant region of the Fab was omitted to reduce computational time.

Protein Production

eOD and gp120 Production. Unless otherwise stated, gp120 and eOD proteins were produced in FreeStyleTM 293F (Invitrogen) suspension cultures by transient transfection using 293Fectin (Invitrogen) of a pHLSec plasmid containing either mammalian codon-

optimized eOD or gp120 with a C-terminal His_{6x} affinity tag. Protein was harvested from the supernatant after 96 h and purified by affinity chromatography with a HIS-TRAP column (GE) followed by Superdex 75 (for eOD) or Superdex 200 (for gp120) size exclusion chromatography (GE Healthcare) using an AKTA Express system (GE Healthcare).

***E. coli* Fab Production.** DNA segments encoding the Fab heavy and light chains containing His_{6x} and AviTag affinity tags were synthesized. DNA was codon optimized for *E. coli* expression and RNA structure, subcloned into pBAD-DEST49 Gateway® Destination Vector (Invitrogen) and transformed into TOP10 E. coli (Invitrogen). Single colonies were grown overnight at 37 °C in 10 mL Luria Broth (LB) plus Ampicillin (100 mg/mL). The starter cultures were expanded into 1 L of Terrific Broth (Fisher Scientific) plus Ampicillin and incubated at 37°C; when cells reached an OD600 of 1.0, Arabinose was added to the cultures to a final concentration of 0.05% to induce protein expression and the cells were then incubated overnight at 18°C. Cultures were pelleted and re-suspended in Start Buffer (20 mM imidazole, 500 mM sodium chloride, 20 mM sodium phosphate), 1 tablet of protease inhibitor (Novagen). The cell suspension was thawed and lysed by the addition of 10x Bugbuster (Novagen) to a final concentration of 1X, 20 µL of Benzonase Nucleases (Novagen) and 1.0 µL of rLysozyme (Novagen) and gently rocked on a rotating mixer for 60 minutes. Lysed cells were pelleted and the supernatant was filtered through a 0.22 µm filter (Millipore). Supernatants were then passed over a 5 mL HisTrap FF crude affinity column (GE Healthcare). The resin was washed with 50 mL Wash Buffer (50 mM imidazole, 500 mM sodium chloride and 20 mM sodium phosphate, pH 7.4) and eluted with 20 mL of Elution Buffer (500 mM imidazole, 500 mM sodium chloride and 20 mM sodium phosphate, pH 7.4). Fractions containing the construct of interest were combined and further purified by affinity chromatography with a KappaSelect resin (GE Healthcare) at room temperature and eluted with 0.1 M glycine (pH 3.0) and immediately neutralized by adding an equal amount of PBS + 0.1M Tris (pH 8.0). The sample was then dialyzed into PBS (pH 7.4) overnight. Collected

fractions were analyzed on a 12% SDS denaturing gel (Invitrogen) and positive fractions were combined and concentrated by ultrafiltration (Vivaspin, Bioexpress).

Mammalian Fab Production. Recombinant Fabs were produced in FreeStyleTM 293F (Invitrogen) suspension cultures by co-transfection of pHLSec plasmids containing expression constructs for light chain and Fab heavy chain using 293Fectin (Invitrogen). Protein was harvested after 96 h and the supernatant was concentrated to <200 mL by tangential flow concentration. Fabs were purified by Kappa Select (GE Healthcare) affinity chromatography followed by size exclusion chromatography (SEC) using Superdex 200 (GE Healthcare) in HBS.

Mammalian IgG Production. IgG Ab variants were produced in FreeStyleTM 293F (Invitrogen) suspension cultures by transfection of pFUSEss expression vectors (Invivogen) containing the light chain and heavy chain using 293Fectin (Invitrogen). Supernatants were harvested 96 h after transfection and Abs were purified using ProteinA SepharoseTM (GE Healthcare) and dialyzed overnight into PBS (0.01 M sodium phosphate, pH 7.4, 0.137 M sodium chloride).

Affinity Determination

Surface Plasmon Resonance (SPR). All binding experiments were carried out on a Biacore 2000 instrument (GE Healthcare) at 25°C with HBSEP+BSA (0.01 M HEPES pH 7.4, 0.15 M sodium chloride, 3 mM EDTA and 0.005% (v/v) Surfactant P20) (GE Healthcare) + 1mg/mL BSA as running buffer. For binding analysis, ~300 response units (RUs) of IgG were captured on a CM5 sensor chip containing 8000-9000 RUs of amine-linked mouse anti-human IgG (Human Ab Capture kit, GE Healthcare). Samples of different protein concentrations were injected in duplicates over this surface at a flow rate of 50 µL/min. If necessary, surface regeneration was performed with two 60 s injections of 3 M magnesium

chloride at a flow rate of 10μL/min. One flow cell contained anti-human IgG only and its interaction with the analyte was used as reference.

Data preparation and analysis were performed using Scrubber 2.0 (BioLogic Software). For kinetic analysis, biosensor data were globally fit to a mass transport limited simple bimolecular binding model:

$$A_0 \underset{k_m}{\overset{k_m}{\rightleftharpoons}} A + B \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} AB$$

where A_0 represents injected analyte. For equilibrium analysis, each data set was fitted to a single site interaction model:

$$R_{eq} = \frac{R_{max} * K_1 * C_1}{K_1 * C_1 + 1}$$

where R_{eq} is the response value at equilibrium, C_A is the concentration of the analyte, R_{max} is the maximum response obtained when all binding sites are occupied by the analyte and K_D is the dissociation constant.

Preliminary Screen of gp120s. Binding for VRC01 and GL-PGV04 Abs was measured by SPR for YU2, HXB2, Bal, JRFL, JRCSF and DU179 gp120. BaL was the only construct that showed detectable binding to the GL of VRC01, although it did not show binding to the GL of PGV04. BaL core gp120 (based on the sequence http://www.ncbi.nlm.nih.gov/protein/AFJ93245.1) and a D368R knockout variant were displayed on the surface of yeast (*19*) to confirm specificity for the GL-VRC01 Ab. Surprisingly, we were unable to detect binding to either GL Ab on the surface of yeast, possibly suggesting a false positive in the original SPR binding assays. Nonetheless, the BaL sequence was used as one of our starting backgrounds.

Comprehensive Screening of HIV Env interactions with GL VH1-2*02 Abs. Ninety-six-well ELISA plates were coated overnight at 4°C with 50 uL PBS containing 5 ng/uL of sheep D7324 anti-gp120 antibody (Aalto Bio reagents). The wells were washed four times with PBS containing 0.05% Tween-20 and blocked with 3% BSA at room temperature for 1 h. Virus supernatants from 63 different isolates were harvested after transfection in 293T cells and were lysed using a final concentration of 1% NP-40 and 50 uL of lysed virus were added to the wells. After incubating at 37°C for 2 h, the wells were washed four times with PBS containing 0.05% Tween-20. Serial dilutions of mAbs were then added to the wells, and the plates were incubated at room temperature for 1 h. After washing four times, goat anti-human IgG $F(ab')_2$ conjugated to alkaline phosphatase (Pierce) was diluted 1:1000 in PBS containing 1% BSA and 0.025% Tween-20 and added to the wells. The plate was incubated at room temperature for 1 h, washed four times, and the plate was developed by adding 50 uL of alkaline phosphatase substrate (Sigma) to 5 mL alkaline phosphatase staining buffer (pH 9.8), according to the manufacturer's instructions. The optical density at 405 nm was read on a microplate reader (Molecular Devices).

Engineering GL Immunogens

Computational Design. An homology model of the GL-VRC01 was aligned onto the coordinates for the mature VRC01/gp120 complex (PDBID: 3NGB) using a backbone alignment to the mature Ab. Subsequently, a crystal structure of unliganded Gl-VRC01 confirmed the initial homology models. An evident clash between the GL-VRC01 CDRL1 insert and the glycan at position N276 was observed. The glycan was, therefore, removed with an N276D mutation and the redocked complex was minimized to reduce side-chain clashes.

Computational design to improve affinity was carried out using RosettaScripts (*34*) to call sequential modeling tasks in the Rosetta software suite. Initially, small variations in the rigid-body orientation were generated using RosettaDock (*35*) followed by RosettaBackrub (*36*) to sample slight conformations in the protein backbone. The initial modeling was done using low resolution side-chain representations (centroid mode) with the intention of creating diversity in the starting structure. Since the computational interface design was being carried out with a docked homology model rather than a high-resolution structure, confidence in the input structure was limited. Following initial perturbation of the modeled complex, sequence design was performed using RosettaDesign with full atom representation. All side chains within 8 Å of the interface on both partners were allowed to sample alternative rotameric conformations, and interface mutations were sampled on gp120 to minimize the energy of the complex. Five rounds of interface optimization were carried out to generate sufficient diversity. Because small modifications in backbone position can cause significant bias in side-chain design, a backbone conformational variant was generated prior to each design step, by passing the pose through RosettaBackrub followed by RosettaDesign. After the pose was passed through five iterations of backrub followed by design, a final output structure was generated.

This procedure was repeated ~100,000 times with each run producing a unique model. Output decoys were filtered based on Rosetta calculated total score and binding energy (Ecomplex –Epartner1 – Epartner2) and, then loosely filtered on unsatisfied polar residues and RMSD from the starting structure. Sequences from the top ~100 decoys were aligned, and the predicted mutations were manually inspected. We specifically selected mutations that made backbone contacts with the Ab or targeted side chains that were identical between both the GL and the mature Ab. Mutations that passed manual inspection were used to create directed libraries of gp120 mutants predicted to have increased affinity for the GL-VRC01.

Surface Display and Selection Overview. Positions on gp120 identified by Rosetta to improve GL Ab binding were screened experimentally using yeast surface display and fluorescence-activated cell sorting (FACS). Directed libraries were constructed using PCR assembly with partially degenerate primers (*37*) encoding the gp120 or eOD of interest. Mutations at positions that were computationally predicted to be beneficial for GL binding, as well as the wild-type amino acid were used to generate a library using degenerate codons so that all possible combinations of desired amino acids would be sampled at that position. The PCR assembly product was gel purified and cloned into pCTCON2 for yeast surface display via homologous recombination in yeast (*38*). gp120 variants were expressed on the surface of EBY100 as a fusion protein between Aga2p and a C-Myc tag (*19*). Biotinylated (Biotin Protein Ligase, Avidity) GL-VRC01 Ab was pre-incubated with phycoerythrin-conjugated streptavidin (SAPE) (Invitrogen) to create tetrameric complexes to maximize avidity. SAPE/Ab complexes were purified by Superdex 200 size exclusion chromatography (GE Healthcare) using an AKTA Avant system (GE Healthcare).

The yeast library was induced at an optical density of $1x10^7$ for ~12 h in C-Trp-Ura + 2% galactose dropout media at 30°C. Induced cells were pelleted, washed and suspended in PBSA buffer (0.01 M sodium phosphate, pH 7.4, 0.137 M sodium chloride, 1 g/L bovine serum albumin) and were duel-labeled with fluorescein isothiocyanate labeled α-cMyc Ab (Immunology Consultants Laboratory) and the SAPE/GL-VRC01 Fab complex at 4°C for 1 h.

Cells were analyzed by FACS on a BD Influx (BD Biosciences) and double positive clones were collected, expanded and re-induced for additional rounds of selection. After modest binding was achieved, the above protocol was modified such that cells were first labeled with biotinylated Fab, washed in PBSA, then labeled with SAPE for fluorescence.

After several rounds of enriching for binding to GL Abs, sequences were recovered. Cells were plated at a density of ~200 clones per plate on C-Trp-Ura + 2% glucose agarose plates, plasmid DNA was extracted and sequenced (Genewiz, La Jolla).

Development of germline-targeting Core gp120. After 3 rounds of sorting using libraries on core BaL [based on http://www.ncbi.nlm.nih.gov/protein/AFJ93245.1] and 93TH057, clones had enriched for the core BaL gp120 library that bound to GL-VRC01, but not to GL-PGV04. No clones enriched from the 93TH057-based library when sorted against either antibody. After 5 rounds of sorting, sequences were recovered from the core BaL library that bound to GL-VRC01. A representative sequence was produced as soluble gp120 (referred to as CoreBal-GT1) in 293F cells, and was found to bind GL-VRC01 (Table 1). As a control, we also produced core gp120 BaL [based on http://www.ncbi.nlm.nih.gov/protein/AFJ93245.1] on which the library was initially based, as well as another quasi-species of BaL [http://www.ncbi.nlm.nih.gov/protein/AAR05834.1] that lacked a glycan at position N276 due to a T278A mutation. The latter protein bound to GL-VRC01 with a K_D of 4.6 μM, however, a D368R knockout mutation only reduced binding to GL-VRC01 to 6.9 μM and improved binding to GL-PGV19 to 10 nM, so we hypothesized that this interaction may not represent specific binding.

Development of germline-targeting eOD. We had separately developed an engineered outer domain lacking the inner domain, V1, V2, and V3 regions, and β20/21, to serve as a minimal antigen for the CD4bs and other epitopes (*17*) and Huang *et al*., in preparation. To take advantage of these properties for germline-targeting, we sought to develop germline-targeting eODs. This was further encouraged during our development of germline-targeting gp120 by our finding that one of the gp120 modifications that enriched was a shortening of β20/21 by 6 amino acids.

Using eOD instead of core gp120, new homology models were generated and the Rosetta interface design protocol was used to predict eOD specific mutations that would improve GL Ab binding. Rosetta-directed libraries were combined with mutations that had enriched from BaL core gp120 library sorts (and including two loops from BaL) and sorted as described above on the surface of yeast for GL-VRC01 and GL-PGV04 binding. Clones that showed binding were sequenced, the resulting protein (eOD-GT1) was expressed, and the K_D for GL-VRC01 binding was measured to be 44 μM by SPR (table S2). Mutations from that clone were modeled, and computational affinity maturation was carried out a second time, with a starting structure containing the previously identified mutations. This second computational library was screened on the surface of yeast as described above and resulted in eOD-GT2, which bound GL-VRC01 with a K_D of 15 μM by SPR (table S2).

Further optimization was carried out on a library containing error prone (EP) PCR-generated random mutations [GeneMorph II Random Mutagenesis Kit, Agilent Technologies] on the DNA recovered from eOD-GT2. The resulting eOD (eOD-GT3) had a 220 nM affinity for GL-VRC01 (table S2). A second round of error prone mutagenesis was conducted on the DNA that produced eOD-GT3, and screening of that error prone library enriched for an L260F mutation as well as a K464N/E mutation. Rather than testing those mutations individually, a larger computational library was constructed that allowed both of those positions to vary among the original and EP mutations as well as additional Rosetta directed mutations. The resulting computational/random mutagenesis library produced eOD-GT4 that bound GL-VRC01 with a K_D of 34 nM (table S2).

During the course of eOD-GT development, other VH1-2 Abs became available, including NIH45-46 and PGV19. GL-NIH45-46 bound to eOD-GT4 with a K_D of 1 μM and GL-PGV19 bound with a K_D of 27 nM as measured by SPR (table S2). We then used a multistate sorting protocol **(**fig. S22**)** to optimize against multiple targets to identify the minimal number of

mutations necessary for binding to diverse GL and mature VH1-2 Abs. A library was generated in which the original residue as well as others that had previously enriched in our sorts was sampled. The resulting library was divided in two and each was screened against GL-VRC01 or GL-NIH45-46. After two sorts, each library was split in half again, and one half was sorted against the opposite Ab while the other half was further sorted on the original Ab. 50 clones from each of the four libraries were sequenced. We found that some of the mutations that enriched for one of the GL Abs did not result in improved binding by the other Ab. From these data, only mutations that were present in all four libraries were selected and used to express eODs for binding assays with both Abs. By SPR, eOD-GT5 bound to GL-VRC01 with an affinity of 530 nM and bound to GL-NIH45-46 with an affinity of 4.8 μM (table S2)**.** We reverted the A281S mutation on loop D, as well as the S365L mutation on the CD4 binding loop, back to their original positions on eOD-base. While these mutations improved GL affinity, they also significantly modified the interface.

To further minimize mutations, a second reversion library was created and that library was divided into seven sub-libraries and sorted for binding to the GL Abs of VRC01, NIH45-46, PGV19, CHA31, PGV04, as well as mature VRC01 and PGV04 Abs. 50 clones from each the GL-VRC01, GL-NIH45-46, GL-PGV19, mature VRC01 and mature PGV04 libraries were sequenced. Common mutations that enriched in all libraries were selected and combined to create eOD-GT6. This final procedure removed four additional mutations while improving affinity for germline antibodies compared to eOD-GT5 (table S2). We suspect the affinity improvements may be due largely to reverting the D457G mutation that had removed a salt bridge between D457 and R469.

eOD-GT6 differed from the starting eOD-base construct by 8 mutations, including the 6 shown in Table 2 along with S465T and N386D, as well as a point deletion at position 356 (del356). The S465T mutation and del356 were relics from the use of two BaL loops in eOD-

GT1. The N386D mutation is discussed in the main text. We tested reversion of the S465T mutation and it had no significant effect on GL-VRC01 affinity (table S7). del356, present in 84% of HIV-1 strains, was considered part of the scaffold.

Point Reversions. Point reversions of the mutations that contributed to improved GL Ab binding in eOD-GT6 were generated and tested for binding by SPR **(**Table 2**)**

Generation of Mouse Abs. The mouse VH gene collection was downloaded from IMGT, translated and aligned. Based on our sorting criteria, we were unable to find any Abs that contained all of the critical contact residues. Specifically, there were no Abs that had both Arg^{H71} and Trp^{H50}, which we believe to be very important for the initial germline binding interaction. We produced the 5 closest mouse VH genes with Arg^{H71} as GL-VRC01 hybrids, in which we replaced the human VH1-2*02 gene of GL-VRC01 with different mouse VH genes **(**fig S18**)**. Binding to these Abs was assessed for eOD-GT6, as well as Core BaL-GT1. No construct showed detectable binding when flowed as analyte at 100 μM concentration.

Generation of Rhesus Abs. To investigate whether a macaque germline repertoire could potentially bind eOD-GT6, we looked for VH1-2-like Ab VH genes in the unannotated genome using both BLAST and Ensembl to look for potential VH genes in the sequenced macaque genome. Several were identified **(**fig. S20**)** and GL-VRC01 VH gene replacements were generated. We assessed binding to eOD-GT6, and the results are described in the main text.

Crystallization and Structure Determination

Crystallography. For crystallization, GL-VRC01 Fab was produced in *E. coli* with a C-terminal His_{6x} tag on the heavy chain. Fab was first purified with a HisTrap column (GE Healthcare) followed by KappaSelect (GE Healthcare). Subsequently, to ensure a

predominant monomeric population, the Fab was reduced and alkylated with DTT and iodoacetamide, as previously described (*39*). The resulting sample was purified by MonoS cation exchange chromatography (Sigma) followed by Superdex 200 size exclusion chromatography (GE Healthcare) using an AKTA Avant system (GE Healthcare). Purified GL-VRC01 Fab in a buffer containing 20 mM Tris, 150 mM NaCl, pH 8.0 was concentrated to 8.5 mg/ml and setup for crystallization trials using the automated CrystalMation robotic system (Rigaku) at the Joint Center for Structural Genomics (www.jcsg.org). Crystals were obtained from a solution containing 20% PEG 3350 w/v, 0.2 M lithium citrate.

For crystallization of eOD-GT6, a minimal glycan (mglyc) construct possessing only two glycosylation sites (N18 and N65, eOD numbering) was designed and subsequently transfected in lab-adapted 293S ($GnTI^{-/-}$) suspension cells. The secreted His_{6x}-tagged protein was isolated from the supernatant by affinity chromatography using HisTrap nickel columns (GE Healthcare). Before purification to homogeneity by Superdex 200 size exclusion chromatography (GE Healthcare), eOD-GT6mglyc was treated with EndoH (NEB) using a protocol similar to that previously described (*40*). The deglycosylated sample was concentrated to ~10 mg/ml and subjected to crystallization trials using the automated Rigaku CrystalMation robotic system at the Joint Center for Structural Genomics (www.jcsg.org). Initial crystal hits were obtained in a solution containing 20% PEG 3350 w/v, 0.2 M magnesium acetate at 4°C.

To obtain complex crystals of GL-VRC01+eOD-GT6, it was first necessary to resurface eOD-GT6 to mutate aspartic acid residues initially introduced to remove sites of glycosylation (NXT/S to DXT/S) to alanine residues (AXT/S). Proteins were expressed and purified as described above. After incubation in molar excess of eOD-GT6, the complex was Endo-H (NEB) treated and subsequently purified to homogeneity by Superdex 200 size exclusion chromatography (GE Healthcare). The deglycosylated sample was concentrated to ~10

mg/ml and subjected to crystallization trials using the automated Rigaku CrystalMation robotic system at the Joint Center for Structural Genomics (www.jcsg.org). Initial crystal hits were obtained in a solution containing 20% PEG 4000 w/v, 0.1 M sodium citrate, pH 5.6, 20% 2-propanol v/v. All crystal were prepared for X-ray diffraction analysis by first cryoprotecting them in the mother liquor supplemented with 20% glycerol followed by fast plunging into liquid nitrogen.

X-ray Data Collection, Data Processing and Structural Determination. Flash-cooled crystals of unliganded GL-VRC01, EndoH-treated eOD-GT6mglyc, and EndoH-treated eOD-T6+GL-VRC01 were subjected to high energy X-ray radiation at the Advanced Photon Source (APS) or Stanford Synchrotron Radiation Lightsource (SSRL) and diffraction images were collected with strategies leading to datasets with high completeness and redundancy. Data processing was performed using XDS (*41*). Statistics for data collection and processing are reported in table S4. The GL-VRC01 Fab crystal structure was solved using VRC01 Fab coordinates from the PDBID 3NGB as a search model for molecular replacement in PHASER (*42*). To solve the unliganded EndoH-treated eOD-GT6mglyc crystal structure, coordinates of eOD from PDBID 3TYG were used as a search model for molecular replacement using PHASER (*42*). Finally, individual components were used as search models for molecular replacement of the EndoH-treated eOD-GT6+GL-VRC01 dataset. For all structures, refinement was performed using a combination of CCP4 (*43*), PHENIX (*44*) and COOT (*45*). Diffraction data from the eOD-GT6mglyc crystals were indexed in monoclinic space group $P2_1$ with unit cell dimensions of a=44.96 Å, b=217.73 Å, c=44.99Å, β=119.97°. These unit cell dimensions are also consistent with hexagonal, C-centered orthorhombic or C-centered monoclinic lattices; however, data processed in those higher-symmetry lattices had very high R-merge values (>40%). Significant peaks (~70% of the origin height) were observed in the self-rotation function for orthogonal non-crystallographic 2-fold axes in the XZ plane, -30° and 60° from Z. Significant pseudo merohedral-twinning was detected from analysis

with phenix.xtriage (*44*). Molecular replacement was successful in placing four copies of eOD-GT6mglyc in the $P2_1$ cell. Amplitude-based twin refinement in Refmac5 (*45*) was applied throughout refinement, and all six pseudo-merohedral twin operators possible for this lattice were tested (with default R_{merge}< 0.44 and twin domains > 0.07 parameters) resulting in final twin operations (twin fraction) of: 1) h, k, l (0.489); 2) h+l, -k, -l (0.408); and 3) l, -k, h (0.102). Refinement statistics for the final models are reported in table S4.

Multimerization and B Cell Activation

Multimeric Particles. Two different eOD-GT6 multimers were created with the objective to assess their ability to activate B cells *in vitro*. eOD trimers were generated by fusing the C-terminus of eOD-GT6 to the N-terminus of a GCN4 trimer (*15*) (PDBID: 1GCM) with a 12 amino-acid GGSGGSGGSGGG linker. The resulting protein was analyzed by SEC-MALS (Wyatt Corporation, Santa Barbara, CA) to confirm trimer formation **(**fig. S15). Self-assembling 60-mer particles were generated by fusing the C-terminus of lumuzine synthase (*16*) (PDBID: 1HQK) to the N-terminus of eOD-GT6 via a GGSGGSGGSGGSGGG linker (fig. S23). Multimerized particles were purified by lectin affinity. The resulting particles were analyzed by SEC-MALS (fig. S15) and Electron Microscopy (EM). YU2 gp140 trimers were provided by Richard Wyatt. These were confirmed to be predominantly trimeric by SEC-MALS (fig. S15) and confirmed to bind with high affinity to mature VRC01 by SPR (data not shown).

Electron microscopy. For EM, a 3 µL aliquot of the multimerized particles (~0.05 mg/mL) was applied for 5 sec onto a carbon coated 400 Cu mesh grid that had been glow discharged at 20 mA for 30 sec, then negatively stained with 2% uranyl formate for 30 sec. Data were collected using a FEI Tecnai F20 electron microscope operating at 120 keV using an electron dose of 30 $e^-/Å^2$ and a magnification of 100,000x. Images were acquired with a Gatan 4k x 4k CCD camera using a nominal defocus range of 500 to 900 nM (fig. S14).

VRC01 B Cell Activation. B cell lines were used as previously reported (*23*). Briefly, cells were suspended at 4 million cells/mL in Advanced DMEM, labeled with 1.5 μM Indo-1 (Invitrogen) for 30 min at 37°C and washed with 2 mM $CaCl_2$ HBSS, followed by another 30 min at 37°C. Aliquots of 2 × 10^6 cells in 0.5 mL were then stimulated at room temperature with BCR ligands. Ca^{2+} signals were recorded for 180 s, measuring the 405/485-nm emission ratio of Indo-1 fluorescence upon UV excitation. Calcium flux analysis was performed on an LSR II cytometer (BD Biosciences). BCR ligand concentrations were tested at 1 μM, 100 nM, 10 nM and 1 nM. Kinetic analyses were performed using FlowJo (Tree Star).

NIH45-46 B Cell Activation. B cell lines were used as previously reported (*24*). Briefly, the DG-75 human Burkitt's lymphoma (ATCC # CRL-2625) cell line was maintained in RPMI-1640 supplemented with 10% FBS. For electroporation, 2 X10^6 DG-75 cells were suspended in 100 μl of cell line nucleofector solution V (Lonza, Cologne Germany), containing 5 μg of BCR-expressing plasmids and electroporated with the Amaxa Nucleofector II, program O-006 (Lonza, Cologne Germany). At 24 h post-transfection, cells were loaded with Fluo-4 Direct calcium indicator (Invitrogen, Carlsbad CA), in RPMI-1640 medium containing 10% FBS at 37 °C for 45 min. Cells were pelleted and stained with APC-conjugated mouse monoclonal anti-human IgG (BD Pharmingen Cat. # 550931) at a 1/10 dilution in 100 μl of RPMI-1640 with 10% FBS and Fluo-4 Direct for 15 min. The cells were washed with 5 mL in RPMI-1640 containing 10% FBS, pelleted, and resuspended at ~1x10^6 cells/mL in RPMI-1640 and subjected to Ca^{2+} flux analysis at a medium flow rate on an LSR II cytometer (Beckton Dickinson). For all experiments, we gated on B cells expressing comparable numbers of BCRs.

Minimum levels of background fluorescence (Min_{FL}) were determined by averaging the background Fluo-4 absorbance in cells for 30 s, and then activation of exogenous BCRs with various immunogens was determined by monitoring changes in Fluo-4 fluorescence associated with cells expressing the exogenous BCRs (APC positive cells) for 210 s. Ionomycin was added to a final concentration of 6.5 nM for 60 s and maximum Fluo-4 fluorescence (Max_{FL}) was established by averaging changes in Fluo-4 fluorescence recorded during the last 10 s.

The percent of maximum Fluo-4 fluorescence at each time point, t, was determined using the formula: (Fluorescence at t-Min_{FL}) / (Max_{FL}-Min_{FL}) x 100. This analysis was performed on both the BCR positive (anti-IgG-APC positive) and BCR negative cells (anti-IgG-APC negative) simultaneously. The background Fluo-4 fluorescence signal from the BCR negative cells was subtracted from that of the BCR positive population at each time point.

Supplementary Figures:

Fig. S1. Sequence alignment of the heavy chain of the 8 VRC01-class Abs used in this study aligned to their predicted GL VH precursor. Positions of GL-encoded residues important for binding to gp120 are highlighted in red.

```
IGKV1-33*01       DIQMTQSPSSLSASVGDRVTITCQASQDISNYLNWYQQKPGKAPKLLIYDASNLETGVPSRFSGSGSG
12A12             ........................G.G.GSS.Q.............VHG....HR..........FH
3BNC60            .............R...T.......N----G.......RR..........G.K..R...A....RRW.
VRC-CH31          .............L...........RG.GKD.......A.......VS...T..G..........FH

IGKV3-11*01       EIVLTQSPATLSLSPGERATLSCRASQSVSSYLAWYQQKPGQAPRLLIYDASNRATGIPARFSGSGSG
VRC01             ........G........T.II...T..YG.--......R.......V..SG.T..A...D.....RW.
NIH45-46          ..................T.II...T...G.--......R.......V..SG.T..A...D.....RW.

IGKV3-20*01       EIVLTQSPGTLSLSPGERATLSCRASQSVSSSYLAWYQQKPGQAPRLLIYGASSRATGIPDRFSGSGS
PGV04             ..................T.S...T.ASYGH---MT...K....P.K...FAT.K..S.........QF

IGLV2-14*01       QSALTQPASVSGSPGQSITISCTGTSSDVGGYNYVSWYQQHPGKAPKLMIYEVSNRPSGVSNRFSGSK
PGV19             ........................-A...FR.FSS......V..R....L.FS.NR....I.H......
PGV20             .......P...........L....A.T------S.A....YAD...R.IVFDGNK...DI.S.....Q


IGKV1-33*01       TDFTFTISSLQPEDIATYYCQQYDNLP
12A12             .T.SL...G..RD.F...F.AVLEFFGPGTKVEIK
3BNC60            QEYNL..NN.....V...F..V.EFIVPGTRLDLK
VRC-CH31          QN.SL......A..V...F....ETFGQGTKVDIK

IGKV3-11*01       TDFTLTISSLEPEDFAVYYCQQRSNWP
VRC01             P.YN....N..SG..G......YEFFGQGTKVQVDIK
NIH45-46          A.YN.S..N..SG..G......YEFFGQGTKVQVDIK

IGKV3-20*01       GTDFTLTISRLEPEDFAVYYCQQYGSSP
PGV04             .KQY....T.M......R.....LEFFGQGTRLEIR

IGLV2-14*01       SGNTASLTISGLQAEDEADYYCSSYTSSSTL
PGV19             .............I......H.NA.EFFGGGTKVFVLG
PGV20             ..G..........S....Y.H.NAFEFFGGGTKLTVLS
```

Fig. S2. Sequence alignment of the light chain of the 8 VRC01-class Abs used in this study aligned to their predicted GL VL precursor.

```
eOD Base   DTITLPCRPAPPPHCSSNITGLILTRDGGNSNDESEIFRPGGGDMRDIARCQIAGTVVSTQLLLNGSLAEEEVVIRSVDFTDNAKSICVQLNTSVEINC
eOD-GT1    ............................T.D.KT.........................................E..R..................
eOD-GT2    ........................G..I.D.KT.........................................E..R..S...............
eOD-GT3    ........................G..I.D.KT.....S...................................E..R..S...............
eOD-GT4    ........................G..I.D.DT.....S.....................F..............E..R..S...............
eOD-GT5    ........................G..V.D.DT.....S.....................F..................R..................
eOD-GT6    ............................V....T.....S.....................F..................R..................


eOD Base   TGAGHCNISRAKWNNTLKQIASKLREQFGNNKTIIFKQSSGGDPEIVTHSFNCGGEFFYCNSTQLFNSTWFNST
eOD-GT1    ............................-.............F..............D.....D....D..
eOD-GT2    ............................-......S..L.....F..............D.....D....D..
eOD-GT3    ............................-.R....S..L.....F..............D.....D....D..
eOD-GT4    ............................-.R....S..L.....F..............D.....D....D..
eOD-GT5    ............................-.R....S........F..............D.....D....D..
eOD-GT6    ............................-.R.............F..............D.............
```

Fig. S3. Sequence alignment of the eOD-Base and eOD-GT variants.

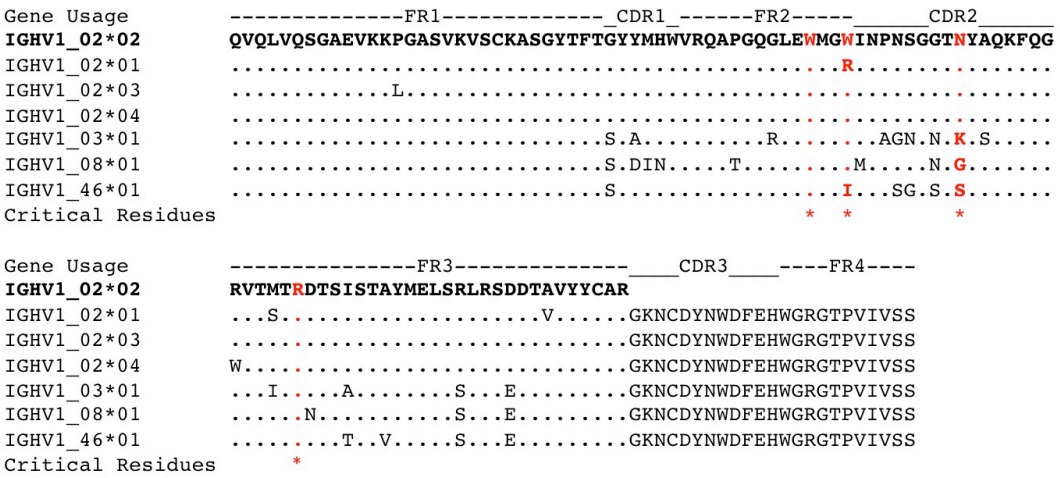


Fig. S4. Sequence alignment of human VH1-2 alleles and closely related human VH genes, with positions of GL encoded residues important for binding to gp120 highlighted in red.

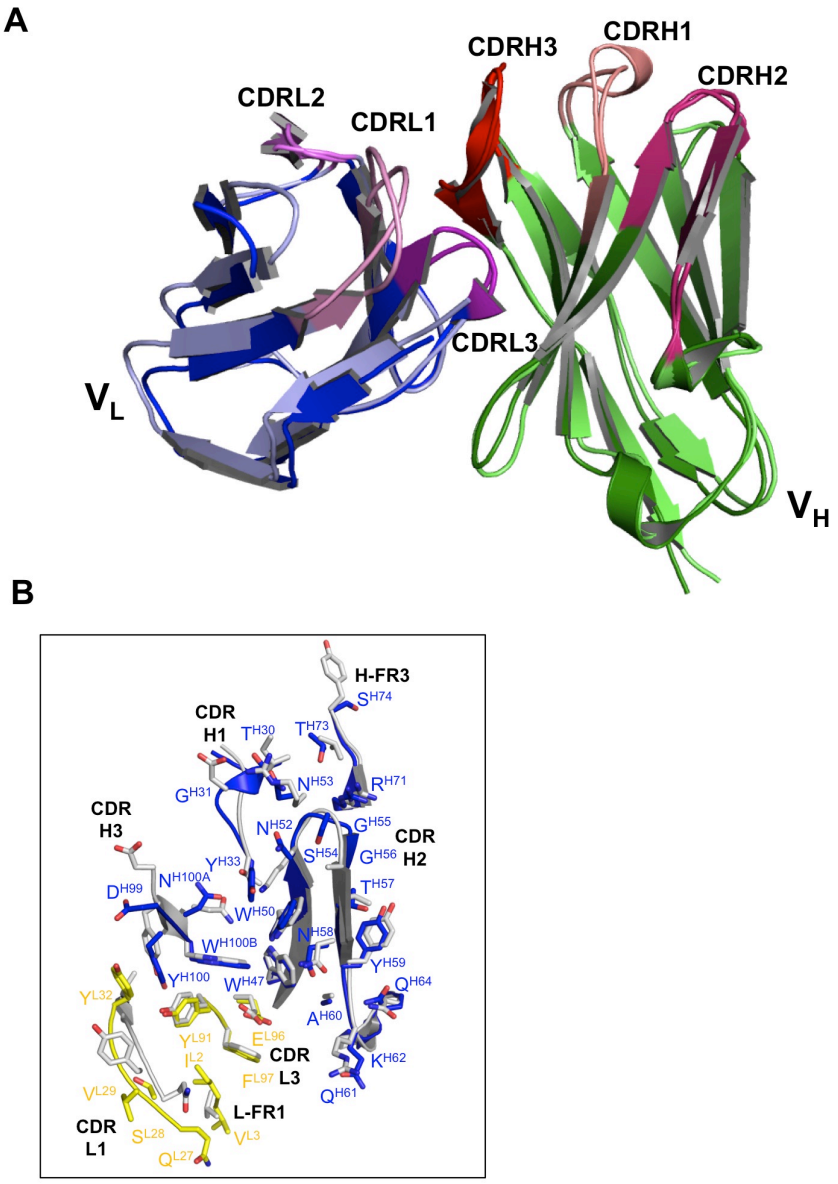


Fig. S5. (A) Comparison between the crystal structures of an unliganded GL-VRC01 Ab (heavy and light chains are colored light green and light blue, respectively) and a gp120-bound mature VRC01 Ab (PDBID: 3NGB, heavy and light chains are colored green and blue, respectively). Only the variable regions are shown for clarity, and structures are rendered as secondary structure cartoons. The CDR loops have been colored differently and are labeled. The overall conformation of the GL and mature VRC01 Fv is similar (core RMSD of 1.2 Å, despite only 66% sequence identity). The elbow angle varies significantly between mature VRC01 (223.1°) and GL-VRC01 (161.4°), but this difference can probably be attributed primarily to crystal packing (*46*). (**B**) Side chains that directly mediate gp120 contacts are mostly structurally conserved between the GL and mature VRC01. GL (heavy and light chain colored blue and yellow, respectively) and mature VRC01 (gray) are rendered as sticks and ribbon.

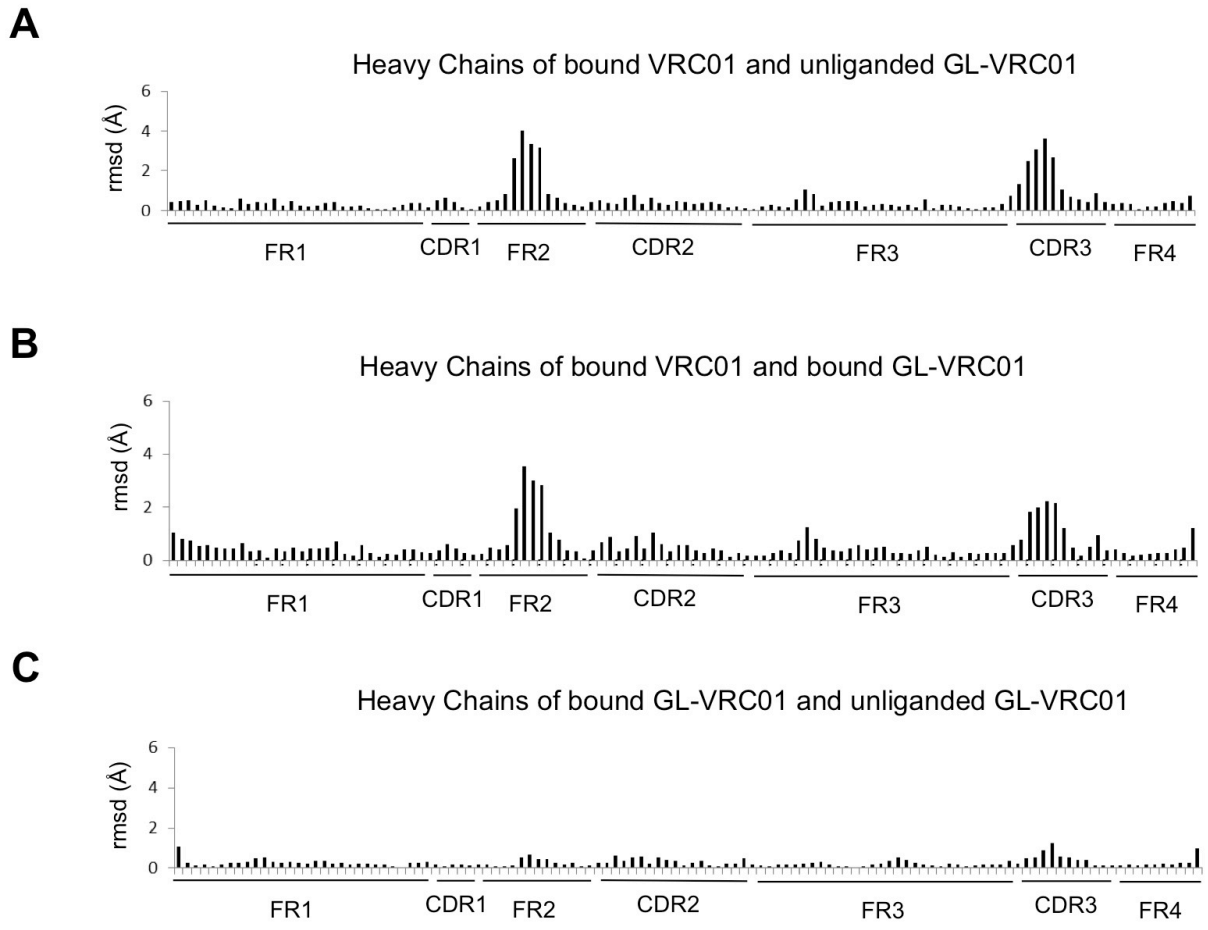


Fig. S6. RMSD comparison between the heavy chains of VRC01 and GL-VRC01 from different crystal structures. The gp120-bound VRC01 crystal structure has PDBID: 3NGB. In (A) and in (B), we observe that the bound-VRC01 varies significantly from GL-VRC01 in the FR2 and CDR3 regions. However, we note that FR2 does not significantly take part in interactions with gp120. Furthermore, analysis of the GL-VRC01 CDR3 loop is limited by the fact that the exact VDJ recombination event in the germline antibody is not predicted in this construct. (C) No large conformational changes are observed in the heavy chain of GL-VRC01 upon binding eOD-GT6. Cα rmsd's were calculated using Chimera (*30*).

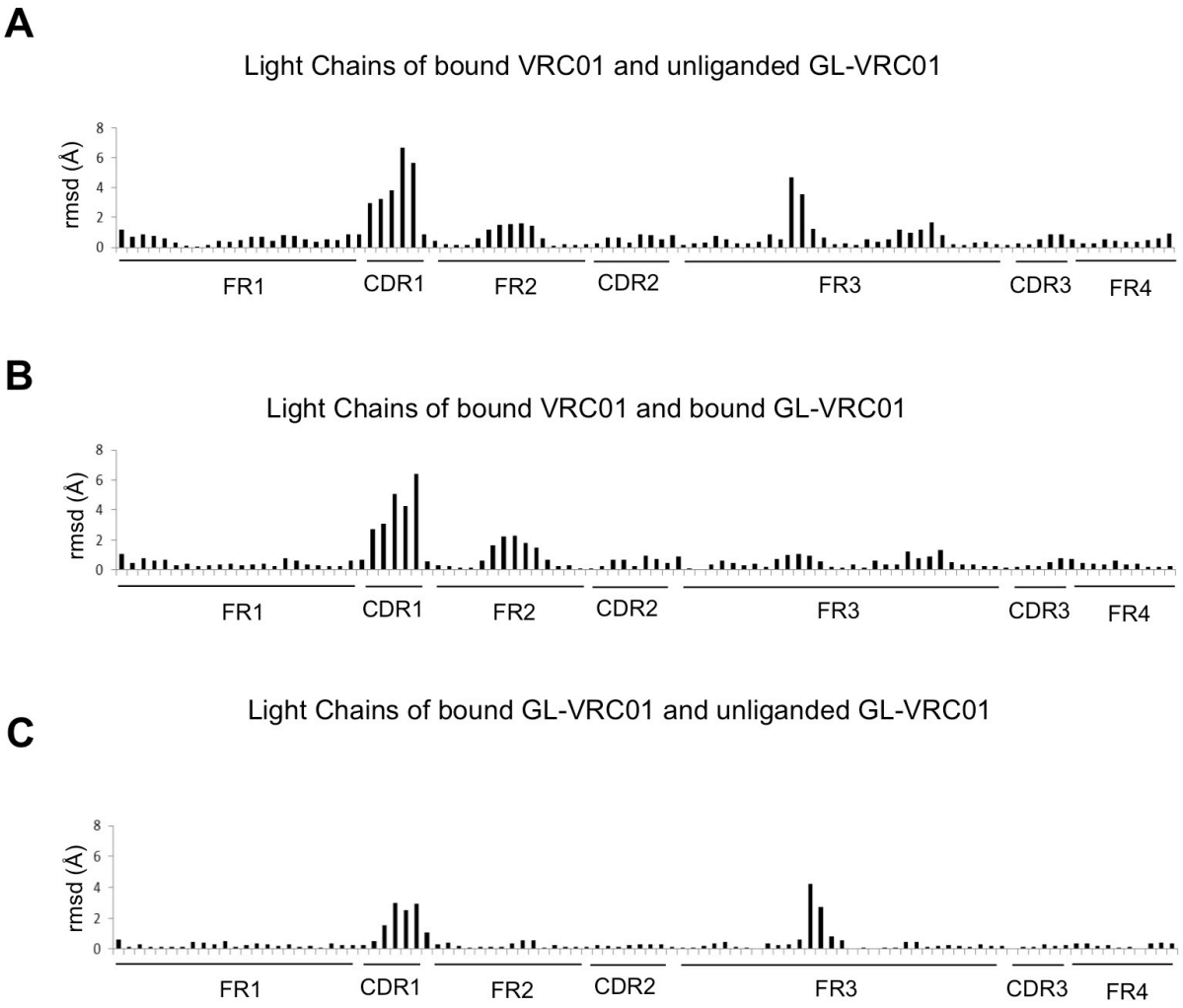


Fig. S7. RMSD comparison between the light chains of VRC01 and GL-VRC01 from different crystal structures. The gp120-bound VRC01 crystal structure has PDBID: 3NGB. In (A) and in (B), we observe that the bound-VRC01 varies significantly from GL-VRC01 in CDR1. A two residue deletion in this region results in conformational changes. (C) Slight conformational changes are observed in the light chain of GL-VRC01 upon binding eOD-GT6, particularly in CDR1 and FR3. Cα rmsd's were calculated using Chimera (*30*).

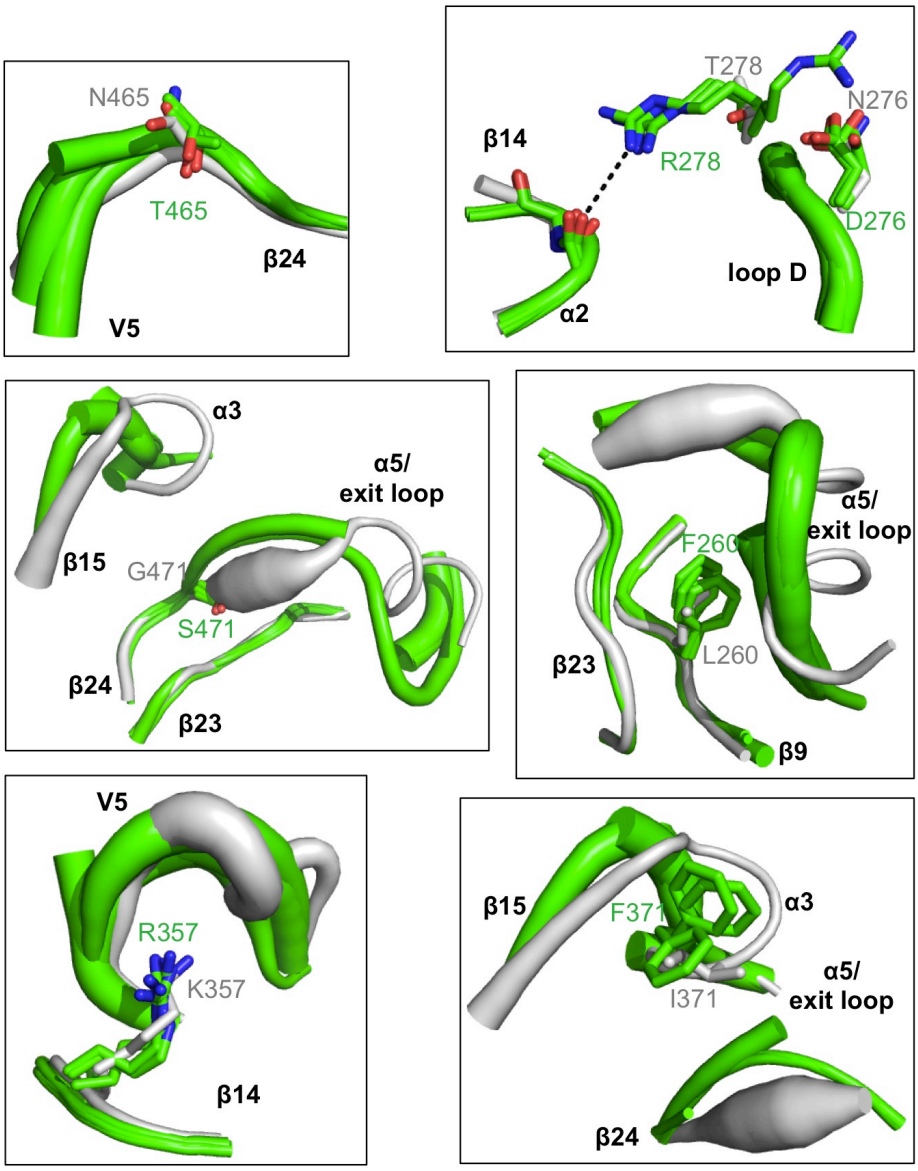


Fig. S8. Detailed view of the mutations in eOD-GT6 that allow binding of GL-VRC-class GL Abs. The crystal structures of unliganded eOD-GT6 (green) and unliganded gp120 core (PDBID: 3TGT, gray) have been superimposed to allow for direct comparison. Also, the four copies present in the asymmetric unit of eOD-GT6 crystal structure have been superimposed and are shown. Structures are rendered according to B-values, with thin and thick lines representing areas of low and high flexibility, respectively. Mutated residues, along with key neighboring residues are shown as sticks. Overall, the unliganded eOD-GT6 crystal structure indicates that key structural features are largely maintained with gp120, despite inner domain truncation and introduction of several mutations in the eOD-GT6 construct.

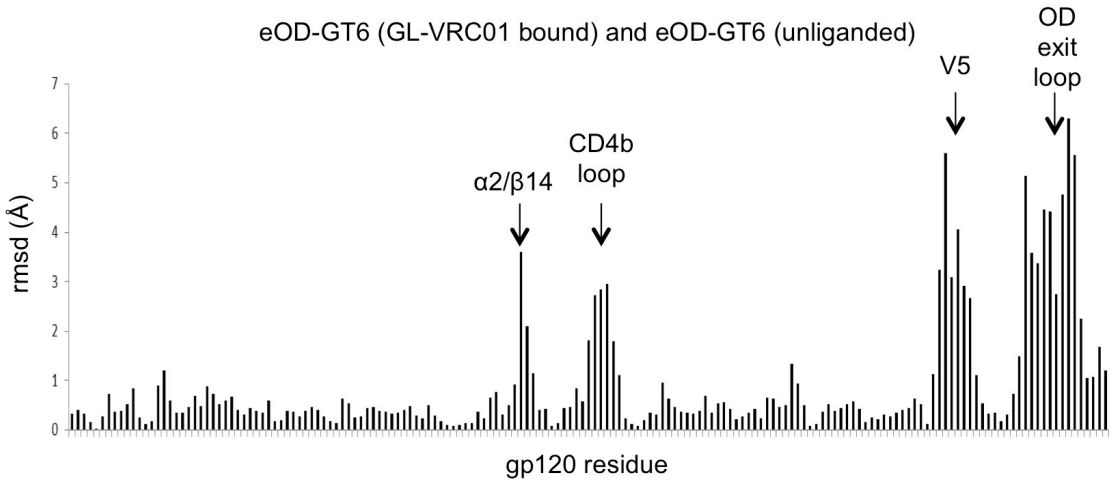


Fig. S9. RMSD comparison between unliganded eOD-GT6 and GL-VRC01 liganded eOD-GT6. Several conformational changes are observed in the eOD-GT6 upon GL-VRC01 binding, particularly in α2/β14, the CD4b loop, V5 and the eOD exit loop. Cα rmsd's were calculated using Chimera (*30*).

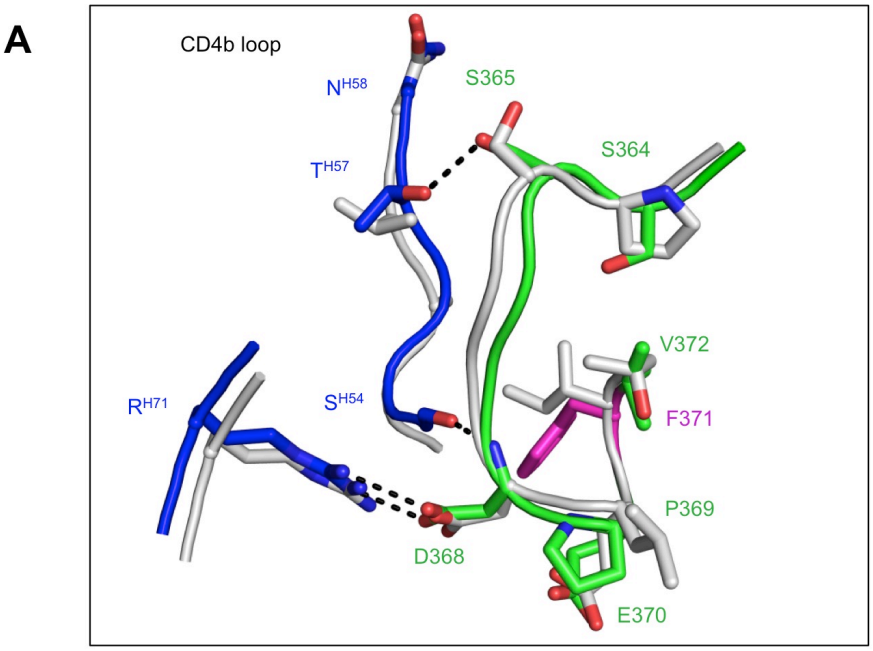


Fig. S10. Comparison between the crystal structures of GL-VRC01+eOD-GT6 and VRC01+gp120 core in the CD4b loop. GL-VRC01 heavy chain is shown in blue. eOD-GT6 is shown in green, with the mutated F371 and V460 in magenta. VRC01+gp120 core components are superposed and shown as gray cartoons. In the CD4b loop, GL-VRC01+eOD-GT6 reveals a mode of recognition nearly identical to that observed in the VRC01+gp120 core (PDBID: 3NGB) structure.

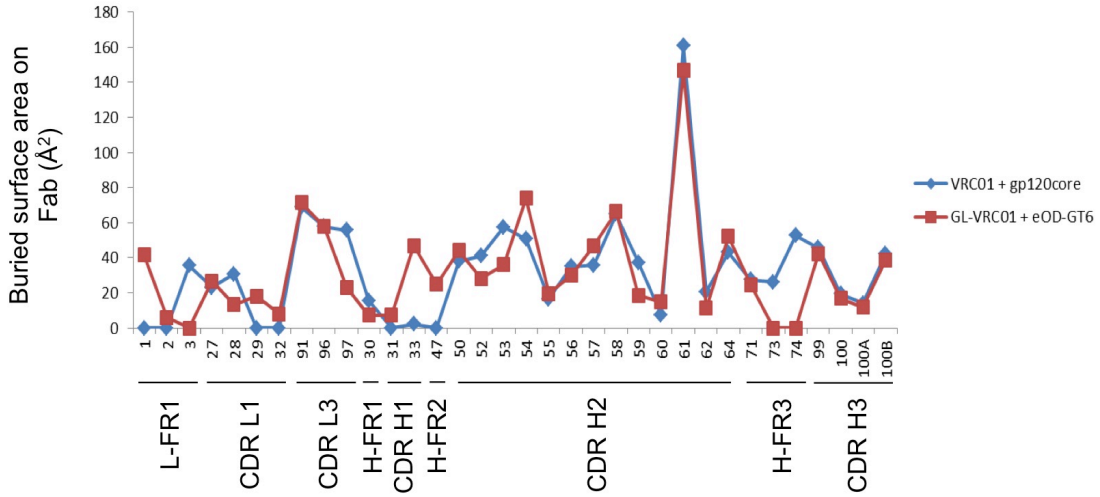


Fig. S11. Graphical rendering of the analysis of the buried surface area and H-bonds on VRC01 and GL-VRC01 in the complexes of VRC01+gp120core and GL-VRC01+eOD-GT6 (Table S5). Interfaces were calculated using PDBePISA (*29*).

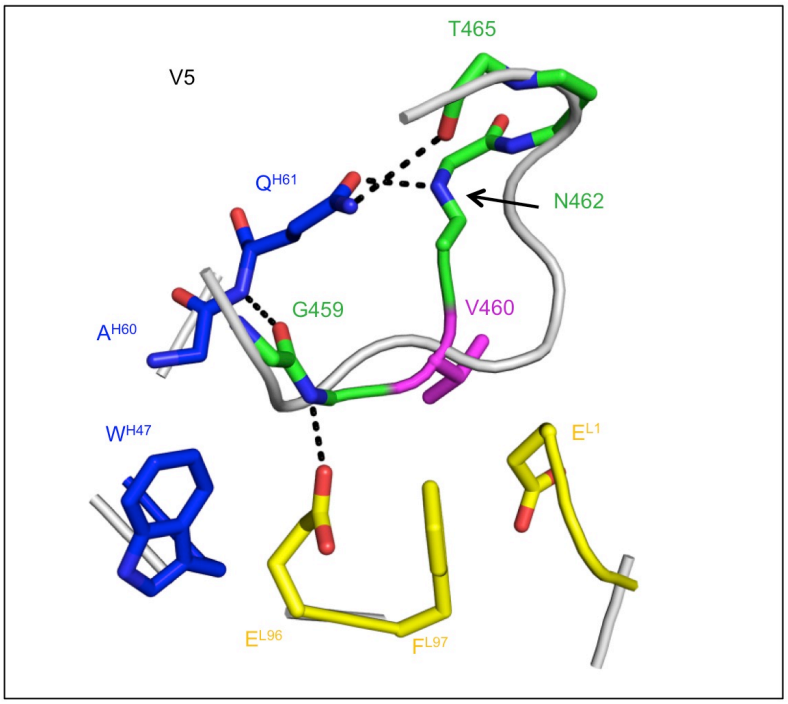
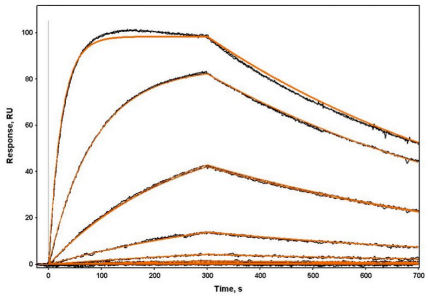


Fig. S12. Comparison between the crystal structures of GL-VRC01+eOD-GT6 and VRC01+gp120 core in the V5 region. GL-VRC01 heavy and light chains are shown in blue and yellow, respectively. eOD-GT6 is shown in green, with the mutated F371 and V460 shown in magenta. VRC01+gp120 core components are superposed and shown as gray cartoons. The GL-VRC01+eOD-GT6 reveals a different mode of recognition of V5 than that observed in the VRC01+gp120 core (PDBID: 3NGB) in that it forms H-bonds with N462 and T465, as opposed to H-bonds with N465 and T467. The conformation of the V5 loop is also significantly different in the two structures.

A

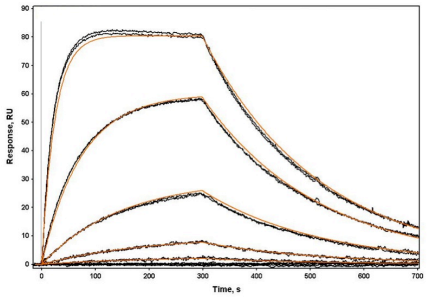
eOD-GT6 R278T (No N276 site)



Injection Concentrations:
4.0 µM
1.0 µM
250 nM
62.5 nM
15.6 nM
3.9 nM
1 nM
250 pM

k_a = 1.0 x 10^4 $M^{-1}s^{-1}$
k_d = 1.6 x 10^{-3} s^{-1}
K_D = 156 nM

eOD-GT6 D276N + R278T (Intact N276 site)



Injection Concentrations:
47.3 µM
11.8 µM
3.0 µM
739 nM
184 nM

k_a = 762 $M^{-1}s^{-1}$
k_d = 4.6 x10^{-3} s^{-1}
K_D = 6.05 µM

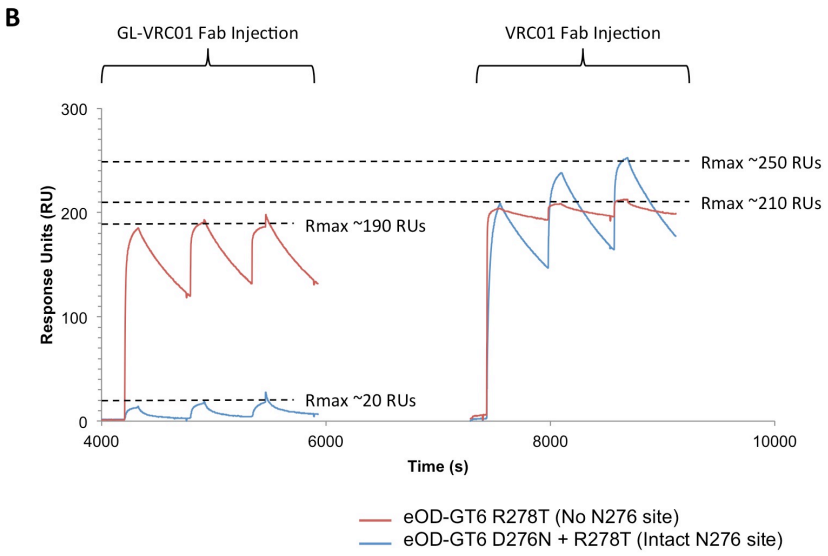


Fig. S13. Surface plasmon resonance experiments show that only ~10% of the eOD-GT6 D276N/R278T double mutant that bound mature VRC01 was able to bind GL-VRC01, suggesting a model accounting for glycan underutilization in which ~10% of the eOD-GT6 double mutant lacked the 276 glycan and had high affinity for GL-VRC01, whereas ~90% utilized the glycan at 276 and had no detectable affinity. (A) SPR traces of GL-VRC01 immobilized while flowing eOD-GT6 with and without the N276 glycan. eOD preps containing the N276 glycan binding site have a significantly reduced on-rate, suggesting that the signal may be due to a small fraction of the sample underutilizing the N276 glycosylation position. (B) eOD-GT6 with and without the glycan immobilized while GL VRC01 and VRC01 Fab are flown as analyte. The significantly different Rmax for the germline antibody for the two constructs suggest that the germline is only capable of binding to a small fraction of eOD, presumably that which lacks the N276 glycan.

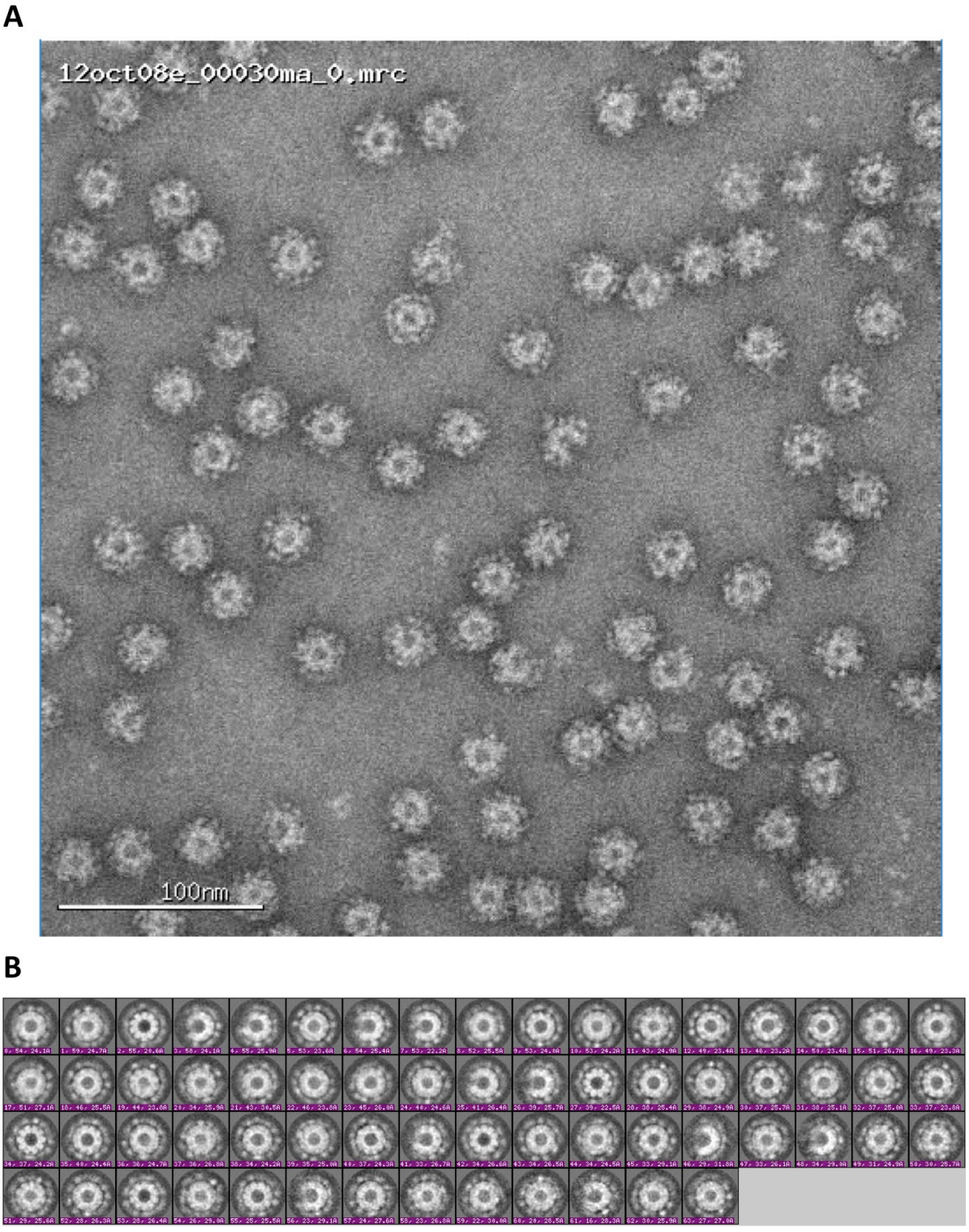


Fig. S14. Negative stain electron microscopy of eOD-GT6 60mer nanoparticles (A) Raw image and (B) 2D-class averages.

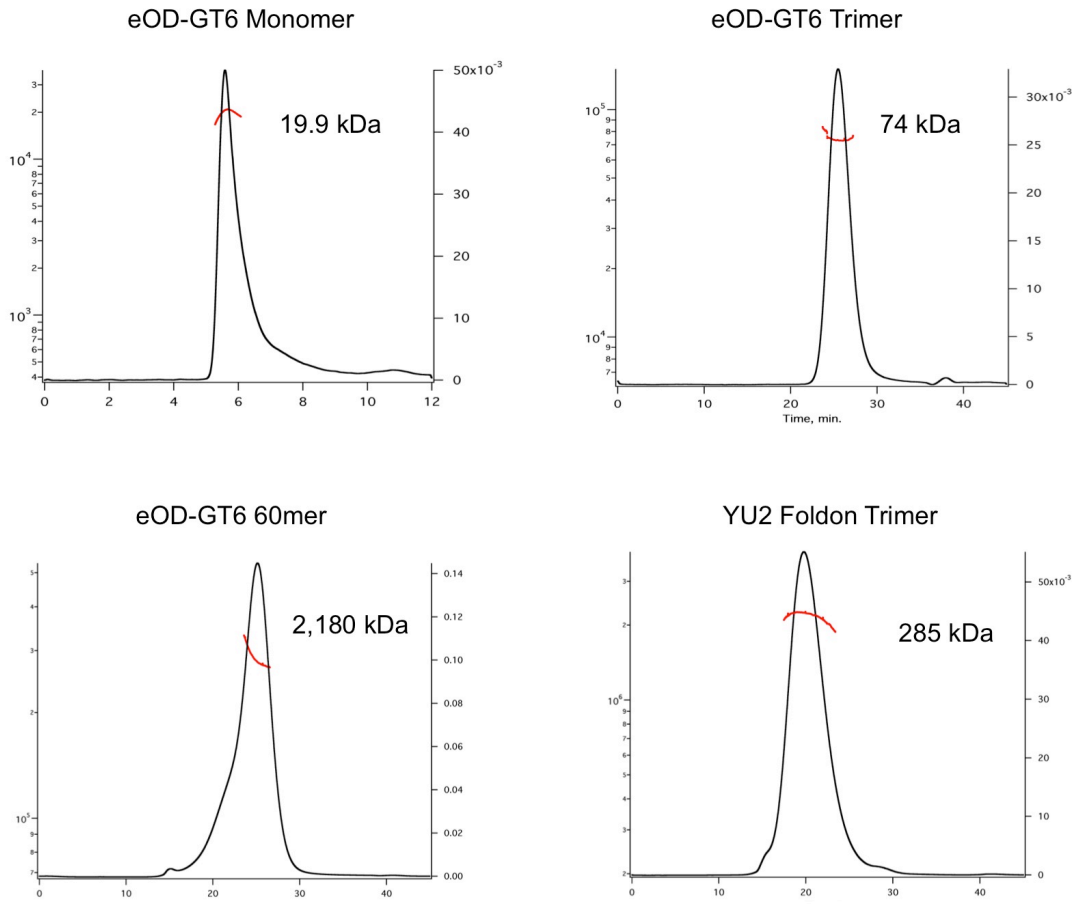


Fig. S15. Purification and characterization of multimeric eOD-GT6 and recombinant soluble YU2 gp140 trimer by size exclusion chromatography coupled in-line with multi-angle light scattering detectors (SEC-MALS). Traces show UV absorbance during the elution and lines under the peak the calculated molar mass of the eluting species. (A) eOD-GT6.0 monomer. Calculated MW: 19.9 kDa, measured MW: 19.9 kDa. (B) eOD-GT6 1gcm trimer. Calculated MW: 71.0, measured MW: 74.0 kDa. (C) eOD-GT6 60mer particle Calculated MW: 2,190 kDa, measured MW: 2,180 kDa.(D) YU2 gp140 trimer. Calculated MW: 237 kDa, measured MW: 285 kDa.

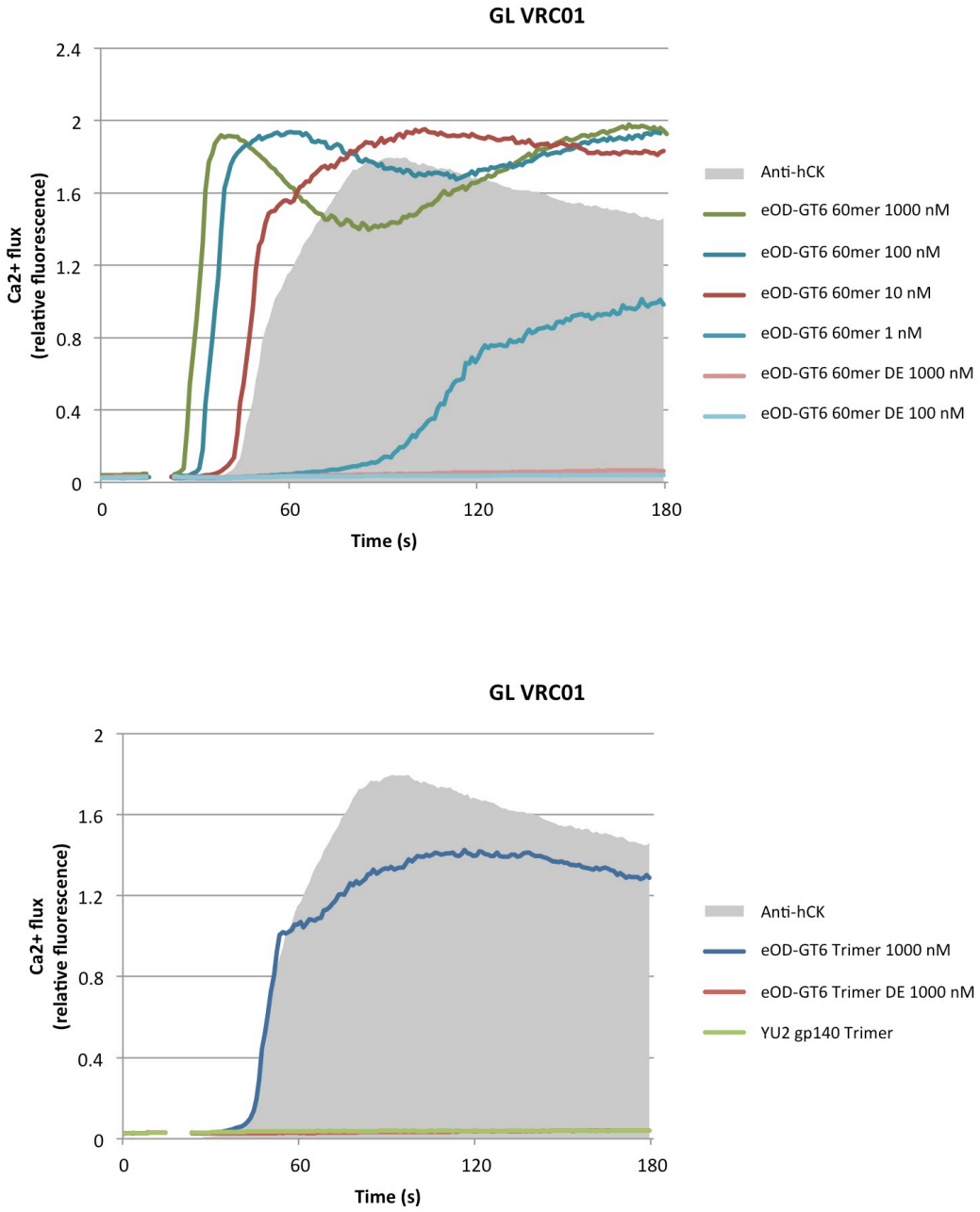


Fig. S16. Ca^{2+} flux activation of B cell lines expressing GL and Mature VRC01 (IgM), 12A12 (IgM and IgG), and NIH45/46 (IgG) with different constructs at varying concentrations. DE = dead epitope, in which a D368R mutation was made in gp120. Absence of activation for this negative control indicates specific activation for the WT construct.

Fig. S16. Cont.

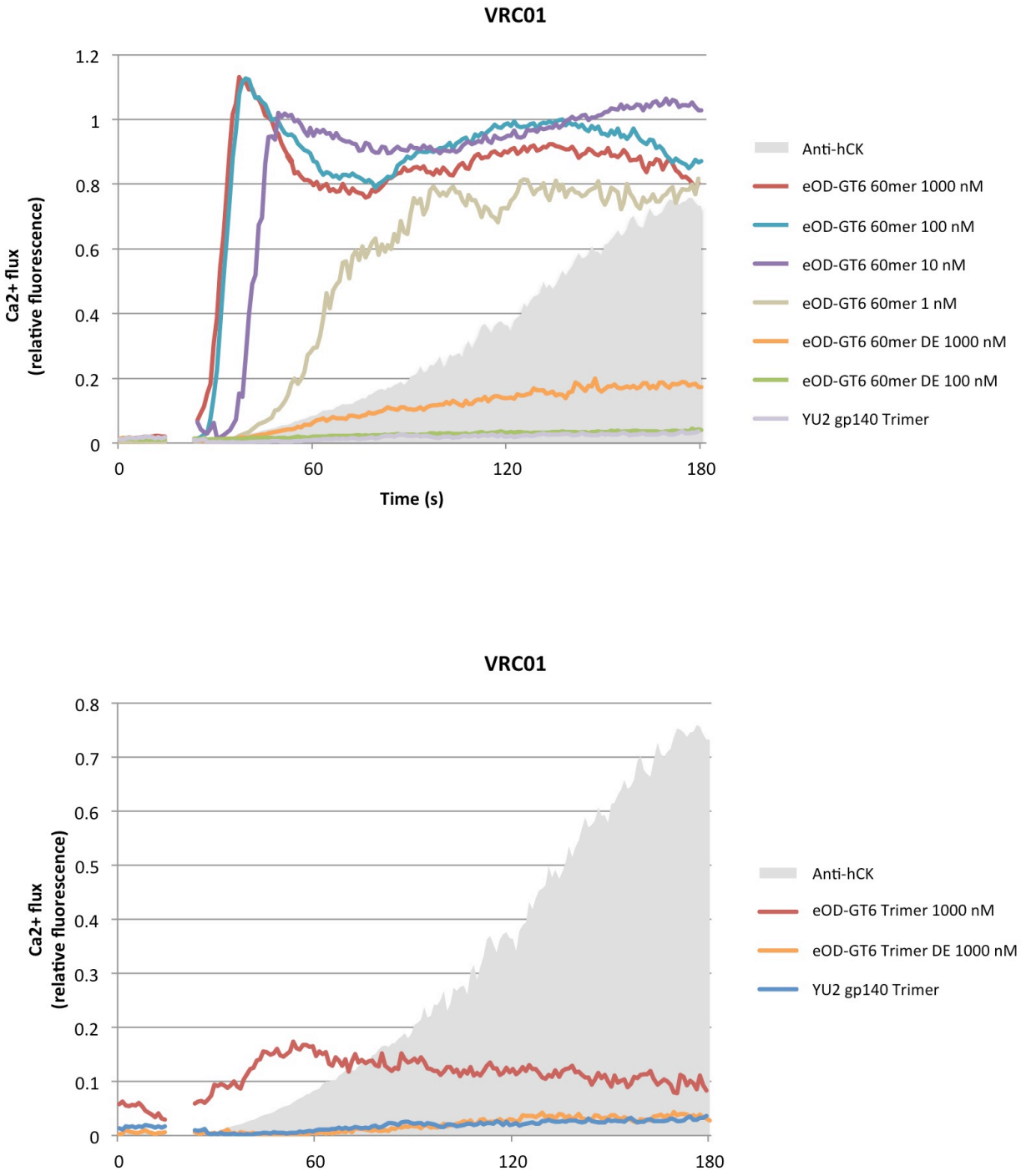
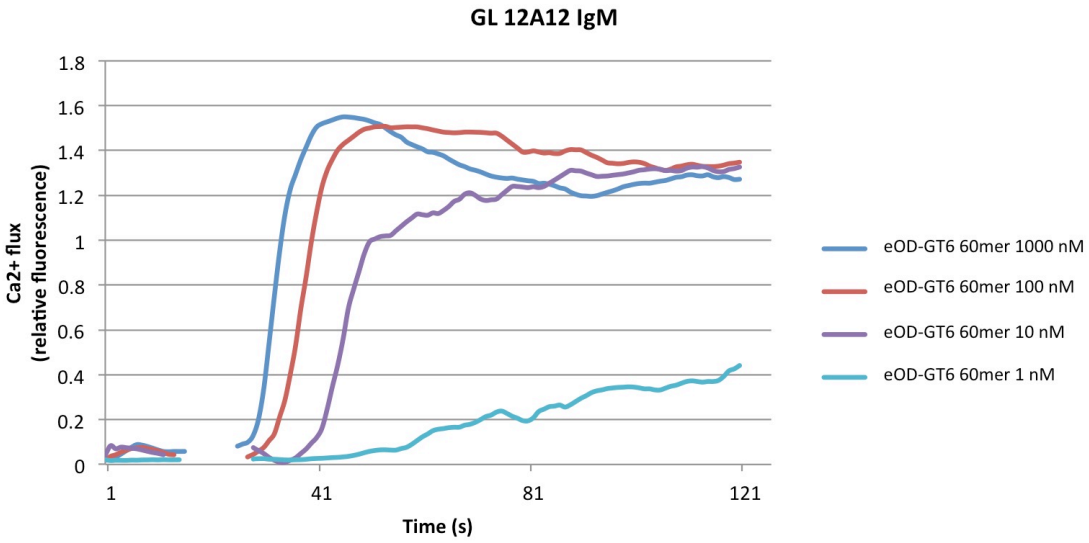

VRC01
Ca2+ flux
(relative fluorescence)
1.2
1
0.8
0.6
0.4
0.2
0
0
60
120
180
Time (s)
Anti-hCK
eOD-GT6 60mer 1000 nM
eOD-GT6 60mer 100 nM
eOD-GT6 60mer 10 nM
eOD-GT6 60mer 1 nM
eOD-GT6 60mer DE 1000 nM
eOD-GT6 60mer DE 100 nM
YU2 gp140 Trimer
VRC01
Ca2+ flux
(relative fluorescence)
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
60
120
180
Anti-hCK
eOD-GT6 Trimer 1000 nM
eOD-GT6 Trimer DE 1000 nM
YU2 gp140 Trimer

Fig. S16. Cont.


GL 12A12 IgM
Ca2+ flux (relative fluorescence)
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
1
41
81
121
Time (s)
eOD-GT6 60mer 1000 nM
eOD-GT6 60mer 100 nM
eOD-GT6 60mer 10 nM
eOD-GT6 60mer 1 nM

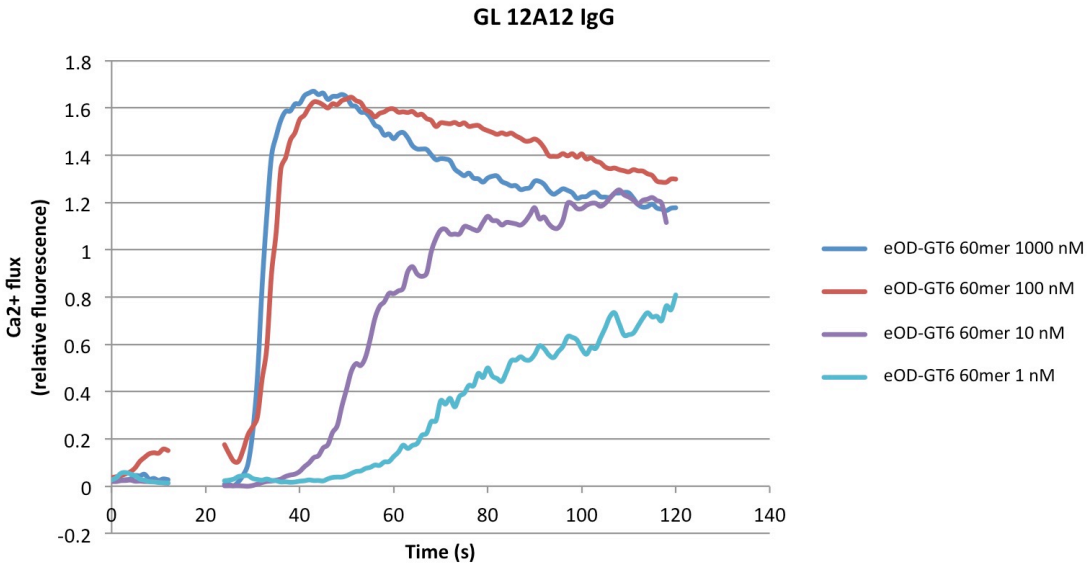

GL 12A12 IgG
Ca2+ flux (relative fluorescence)
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
0
20
40
60
80
100
120
140
Time (s)
eOD-GT6 60mer 1000 nM
eOD-GT6 60mer 100 nM
eOD-GT6 60mer 10 nM
eOD-GT6 60mer 1 nM

Fig. S16. Cont.

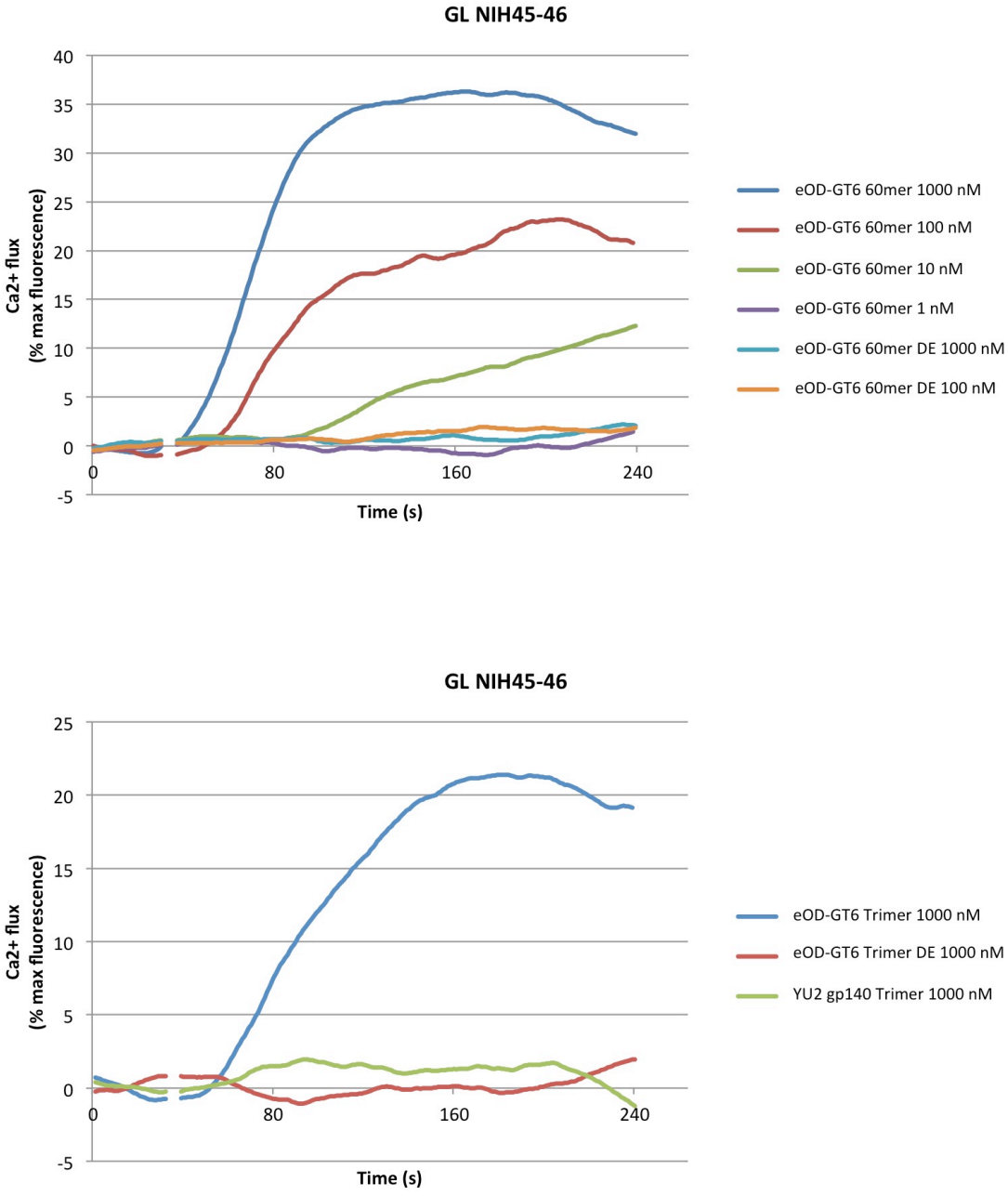
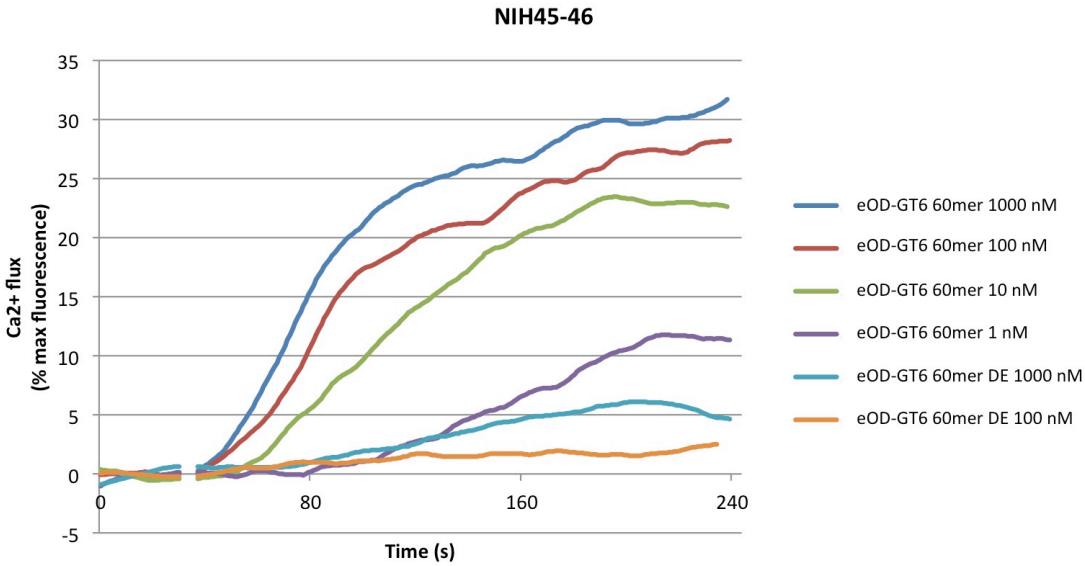
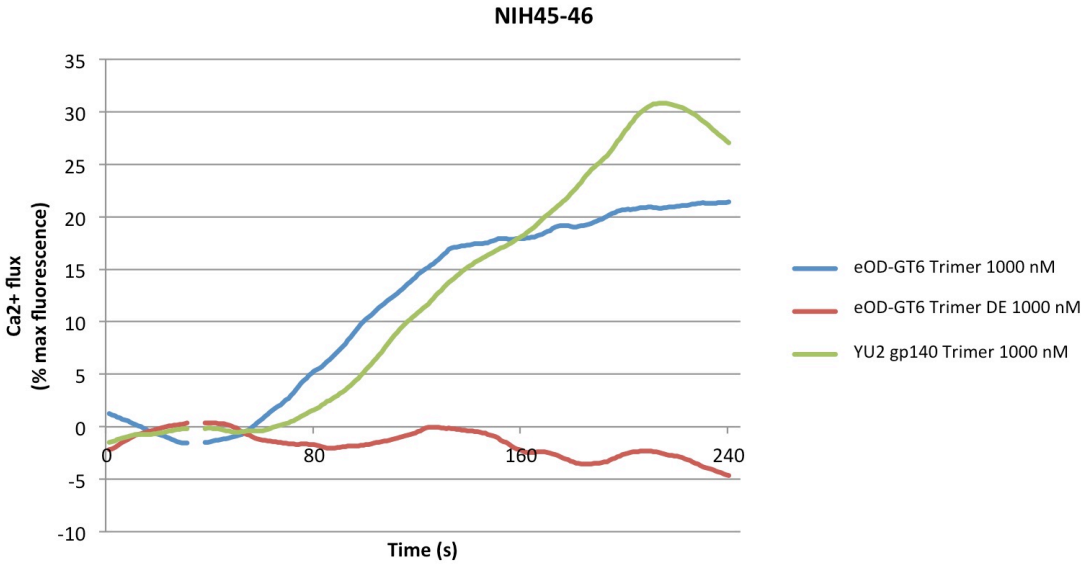

GL NIH45-46
Ca2+ flux
(% max fluorescence)
Time (s)
40
35
30
25
20
15
10
5
0
-5
0
80
160
240
eOD-GT6 60mer 1000 nM
eOD-GT6 60mer 100 nM
eOD-GT6 60mer 10 nM
eOD-GT6 60mer 1 nM
eOD-GT6 60mer DE 1000 nM
eOD-GT6 60mer DE 100 nM
GL NIH45-46
Ca2+ flux
(% max fluorescence)
Time (s)
25
20
15
10
5
0
-5
0
80
160
240
eOD-GT6 Trimer 1000 nM
eOD-GT6 Trimer DE 1000 nM
YU2 gp140 Trimer 1000 nM

Fig. S16. Cont.


NIH45-46
Ca2+ flux
(% max fluorescence)
35
30
25
20
15
10
5
0
-5
0
80
160
240
Time (s)
eOD-GT6 60mer 1000 nM
eOD-GT6 60mer 100 nM
eOD-GT6 60mer 10 nM
eOD-GT6 60mer 1 nM
eOD-GT6 60mer DE 1000 nM
eOD-GT6 60mer DE 100 nM


NIH45-46
Ca2+ flux
(% max fluorescence)
35
30
25
20
15
10
5
0
-5
-10
0
80
160
240
Time (s)
eOD-GT6 Trimer 1000 nM
eOD-GT6 Trimer DE 1000 nM
YU2 gp140 Trimer 1000 nM

A

```
Gene Usage          ----------------FR1-------------_CDR1_------FR2-----______CDR2______
IGHV1-02*02         QVQLV-QSGAEVKKPGASVKVSCKASGYTFTGYYMHWVRQAPGQGLEWMGWINPNSGGTNYAQKFQG
Rabbit VH1-a1       .C.S.EE..GRLVT..TPLTLT.TV..FSLSS.A.S.......K....I.I.S-S..S.Y..SWAK.
Critical Residues                                                  *  *        *

Gene Usage          ----------------FR3--------------____CDR3____----FR4----
IGHV1-02*02         RVTMTRDTSISTAYMELSRLRSDDTAVYYCAR
Rabbit VH1-a1       .F.ISKTST--.VDLKITSPTTE...T.F...GKNCDYNWDFEHWGRGTPVIVSS
Critical Residues       *
```

B

```
Gene Usage          --------------FR1-------------_CDR1_------FR2-----______CDR2______
IGHV1-02*02         QVQLVQSGAEVKKPGASVKVSCKASGYTFTGYYMHWVRQAPGQGLEWMGWINPNSGGTNYAQKFQG
IGHV1S7*01          -Q..KE..GGLV...G.L.LC.....F..SS...C.......K....I.C.YAG..S.H..SWVN.
IGHV1S47*01         .E...E..GGLVQ.EG.LTLT.....FD.SSNA.C.......K.P..IAC.YNGD.S.Y..SWVN.
Critical Residues                                                  *  *        *

Gene Usage          ----------------FR3--------------____CDR3____----FR4----
IGHV1-02*02         RVTMTRDTSISTAYMELSRLRSDDTAVYYCAR
IGHV1S7*01          .F.LS..NAQ..VCLQ.NS.TAA...T.F...GKNCDYNWDFEHWGRGTPVIVSS
IGHV1S47*01         .F.IS.S..LN.VTLQMTS.TAV...T.F...GKNCDYNWDFEHWGRGTPVIVSS
Critical Residues       *
```

Fig. S17. (A) Sequence alignment of the human VH1-2*02 gene with the VH-1 gene reported to be utilized by the majority of rabbit Abs (*47*). This gene only contains 1 of 4 critical residues and is 3 residues shorter, suggesting that the conformation of the backbone might also be different. **(B)** Alignment of the human VH1-2*02 gene with the closest rabbit VH gene. It is unclear if rabbits produce Abs possessing this gene with any useful frequency. Positions of human GL encoded residues important for binding to gp120 are highlighted in red.

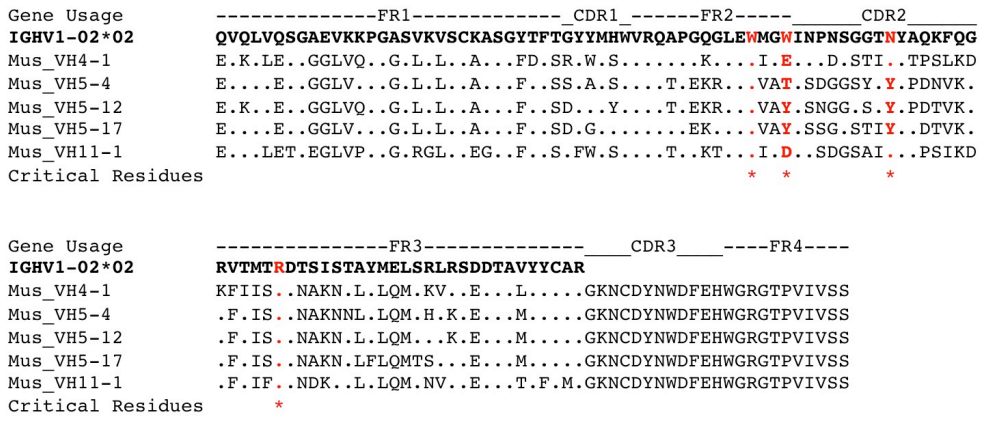


Fig. S18. Sequence alignment of the closest 5 mouse VH genes containing Arg^{H71} to VH1-2*02. Residues colored red represent major incompatibilities with gp120 binding. Positions of human GL encoded residues important for binding to gp120 are highlighted in red stars.

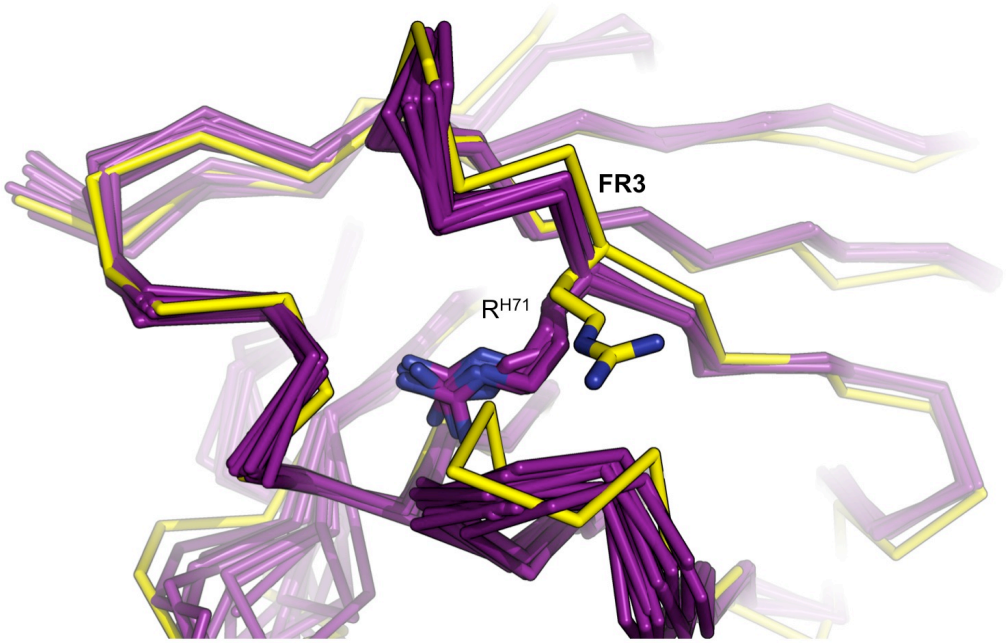


Fig. S19. Structural superposition of 10 Abs from the PDB that originated from the mouse VH5-17*01 gene (PDBIDs 2aab, 2zuq, 3hr5, 2ffd, 1seq, 2qhr, 3ra7, 2vh5, 2xkn, 3s88 and colored purple) and the gp120-bound VRC01 crystal structure (PDBID: 3GNB and colored yellow). There was no density for Arg^{H71} in PDBID: 2AAB, but the remaining 9 structures all have the Arg^{H71} side chain buried into the interior of the Ab structure, while this residue in the VRC01 structure is exposed. In the conformation found in the mouse Abs, Arg^{H71} would not be accessible to form the critical salt bridge with D368 on gp120. Additionally, the conformation of the backbone is significantly modified, further reducing the structural similarities with human VH1-2 Abs.

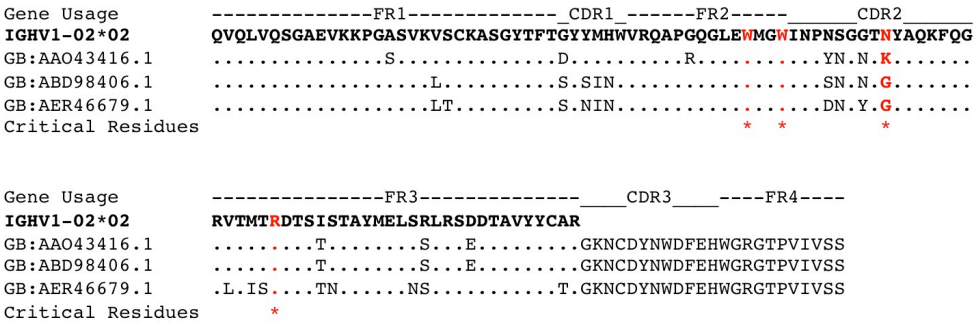


Fig. S20. Sequence alignment of the closest 3 macaque VH genes containing Arg^{H71} to VH1-2*02. While none of the VH genes identified contain all 4 critical contacts, our model suggests that the mutations at position Asn^{H57} do not result in obvious clashes with GP120 and as such, might allow binding to eOD-GT6. Positions of human GL encoded residues important for binding to gp120 are highlighted in red.

```
Gene Usage        --------------FR1-------------_CDR1_------FR2-----_____CDR2_____
IGHV1-02*02       QVQLVQSGAEVKKPGASVKVSCKASGYTFTGYYMHWVRQAPGQGLEWMGWINPNSGGTNYAQKFQG
GL_12A12_1        ............................................................
GL_3BNC60_1       ............................................................
GL_NIH45-46_v1    ............................................................
GL_NIH45-46_v2    ............................................................
GL_PGV04_v1       ............................................................
GL_PGV04_v2       ............................................................
GL_PGV04_v3       ............................................................
GL_PGV19          ............................................................
GL_PGV20          ............................................................
GL_VRC-CH31_v1    ............................................................
GL_VRC-CH31_v2    ............................................................
GL_VRC01_v1       ............................................................
GL_VRC01_v2       ............................................................
GL_VRC01_v3       ............................................................
Critical Residues                                          *  *        *
```

```
Gene Usage        ---------------FR3--------------_______CDR3_______----FR4----
IGHV1-02*02       RVTMTRDTSISTAYMELSRLRSDDTAVYYCAR
GL_12A12_1        ...............................DGSGDD---TSWHF--DPWGQGTLVTVSS
GL_3BNC60_1       ...............................QRSDF------WDF--DLWGRGTLVTVSS
GL_NIH45-46_v1    ...............................GKYCTARDYYNWYF--DLWGRGTLVTVSS
GL_NIH45-46_v2    ...............................GKYSTARDYYNWYF--DLWGRGTLVTVSS
GL_PGV04_v1       ...............................QKFYTGGQ--GWYF--DLWGRGTLVTVSS
GL_PGV04_v2       ...............................QKFARGGQ--GWYF--DLWGRGTLVTVSS
GL_PGV04_v3       ...............................GGGYTGGQ--GWYF--DLWGQGTLVTVSS
GL_PGV19          ...............................MGAARE-----WDF--QYWGQGTRVLVSS
GL_PGV20          ...............................RMRSQD---REWDF--QHWGQGTLVTVSS
GL_VRC-CH31_v1    ...............................GSKRGR---SGWDF--QHWGQGTLVTVSS
GL_VRC-CH31_v2    ...............................GGGRGR---SGWDF--QHWGQGTLVTVSS
GL_VRC01_v1       ...............................GKNCD----YNWDF--QHWGQGTLVTVSS
GL_VRC01_v2       ...............................GKNSD----YNWDF--QHWGQGTLVTVSS
GL_VRC01_v3       ...............................GKYYD----YVWYF—DLWGRGTLVTVSS
Critical Residues      *
```

Fig. S21. Sequence alignment of the GL calculated heavy chain of the VRC01-class Abs used in this study.

```
IGKV1-33*01      DIQMTQSPSSLSASVGDRVTITCQASQDISNYLNWYQQKPGKAPKLLIYDASNLETGVPSRFSGSGSG
GL_12A12         ........................................................................
GL_3BNC60        ........................................................................
GL_VRC-CH31      ........................................................................

IGKV3-11*01      EIVLTQSPATLSLSPGERATLSCRASQSVSSYLAWYQQKPGQAPRLLIYDASNRATGIPARFSGSGSG
GL_NIH45-46_v1   ........................................................................
GL_VRC01_v1      ........................................................................

IGKV3D-15*02     EIVMMQSPATLSVSPGERATLSCRASQSVSSNLAWYQQKPGQAPRLLIYGASTRATGIPARFSGSGSG
GL_NIH45-46_v2   ....T...................................................................
GL_VRC01_v2      ........................................................................

IGKV3-20*01      EIVLTQSPGTLSLSPGERATLSCRASQSVSSSYLAWYQQKPGQAPRLLIYGASSRATGIPDRFSGSGS
GL_PGV04         ........................................................................

IGLV2-14*01      QSALTQPASVSGSPGQSITISCTGTSSDVGGYNYVSWYQQHPGKAPKLMIYEVSNRPSGVSNRFSGSK
GL_PGV19         E.......................................................................
GL_PGV20         E.......................................................................


IGKV1-33*01      TDFTFTISSLQPEDIATYYCQQYDNLP
GL_12A12         ...................AVLEFFGPGTKVEIK
GL_3BNC60        .......................EFIGPGTKVDIK
GL_VRC-CH31      .......................ETFGQGTKLEIK

IGKV3-11*01      TDFTLTISSLEPEDFAVYYCQQRSNWP
GL_NIH45-46_v1   ......................YEFFGQGTKLEIK
GL_VRC01_v1      ......................YEFFGQGTKLEIK

IGKV3D-15*02     TEFTLTISSLQSEDFAVYYCQQYNN*P
GL_NIH45-46_v2   .......................EFFGQGTKLEIK
GL_VRC01_v2      .......................EFFGQGTKLEIK

IGKV3-20*01      GTDFTLTISRLEPEDFAVYYCQQYGSSP
GL_PGV04         .......................LEFFGQGTRLEIK

IGLV2-14*01      SGNTASLTISGLQAEDEADYYCSSYTSSSTL
GL_PGV19         .........................EFFGGGTKVFVLG
GL_PGV20         .........................EFFGGGTKVFVLG
```

Fig. S22. Sequence alignment of the GL calculated light chain of the VRC01-class Abs used in this study.

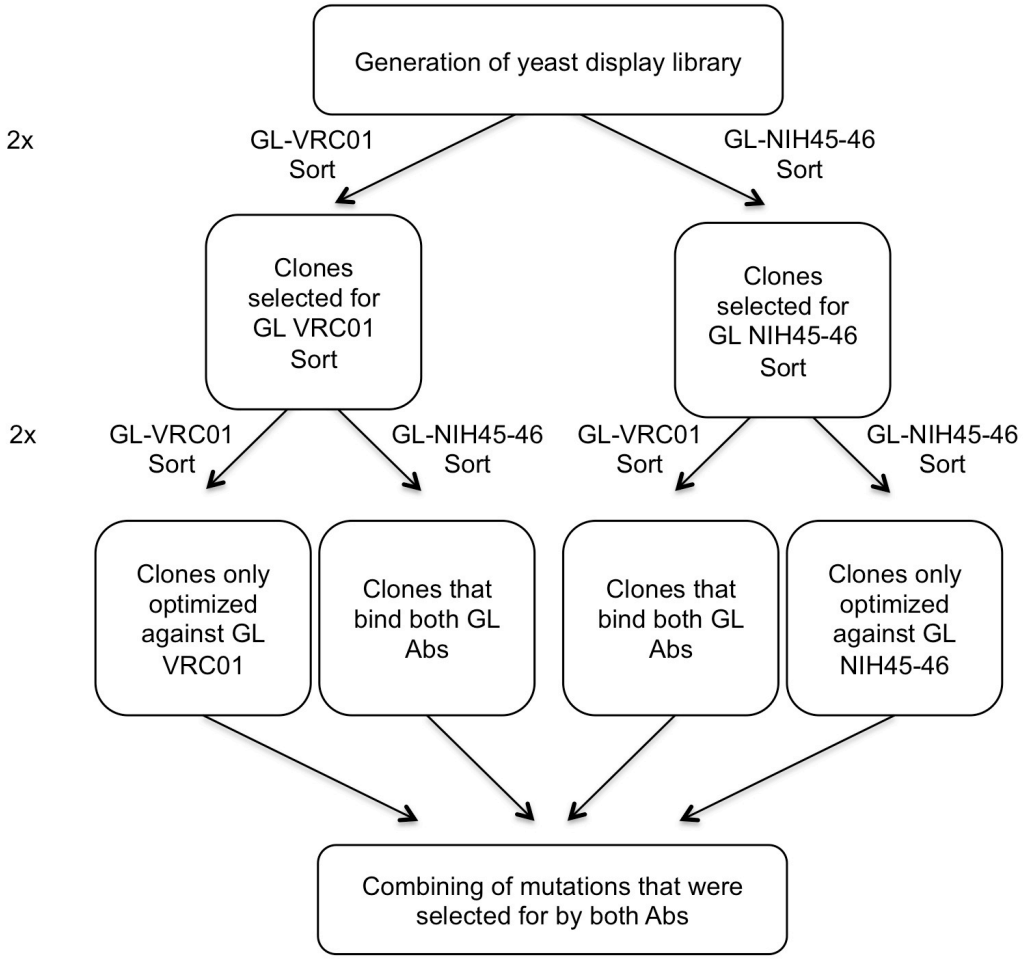


Fig. S23. Flow chart representing the multistate sorting protocol.

MQIYEGKLTAEGLRFGIVASRFNHALVDRLVEGAIDAIVRHGGREEDITLVRVPGSWEIPVAAGE
LARKEDIDAVIAIGVLIRGATPHFDYIASEVSKGLADLSLELRKPITFGVITADTLEQAIERAGT
KHGNKGWEAALSAIEMANLFKSLRGGSGGSGGSGGSGGGDTITLPCRPAPPPHCSSNITGLILTR
DGGVSNDETEIFRPSGGDMRDIARCQIAGTVVSTQLFLNGSLAEEEVVIRSVDFRDNAKSICVQL
NTSVEINCTGAGHCNISRAKWNNTLKQIASKLREQFGNRTIIFKQSSGGDPEFVTHSFNCGGEFF
YCDSTQLFNSTWFNST**

Fig. S24. Sequence of eOD-GT6 60mers. Particle base in red, GS linker in black, eOD-GT6 in blue.

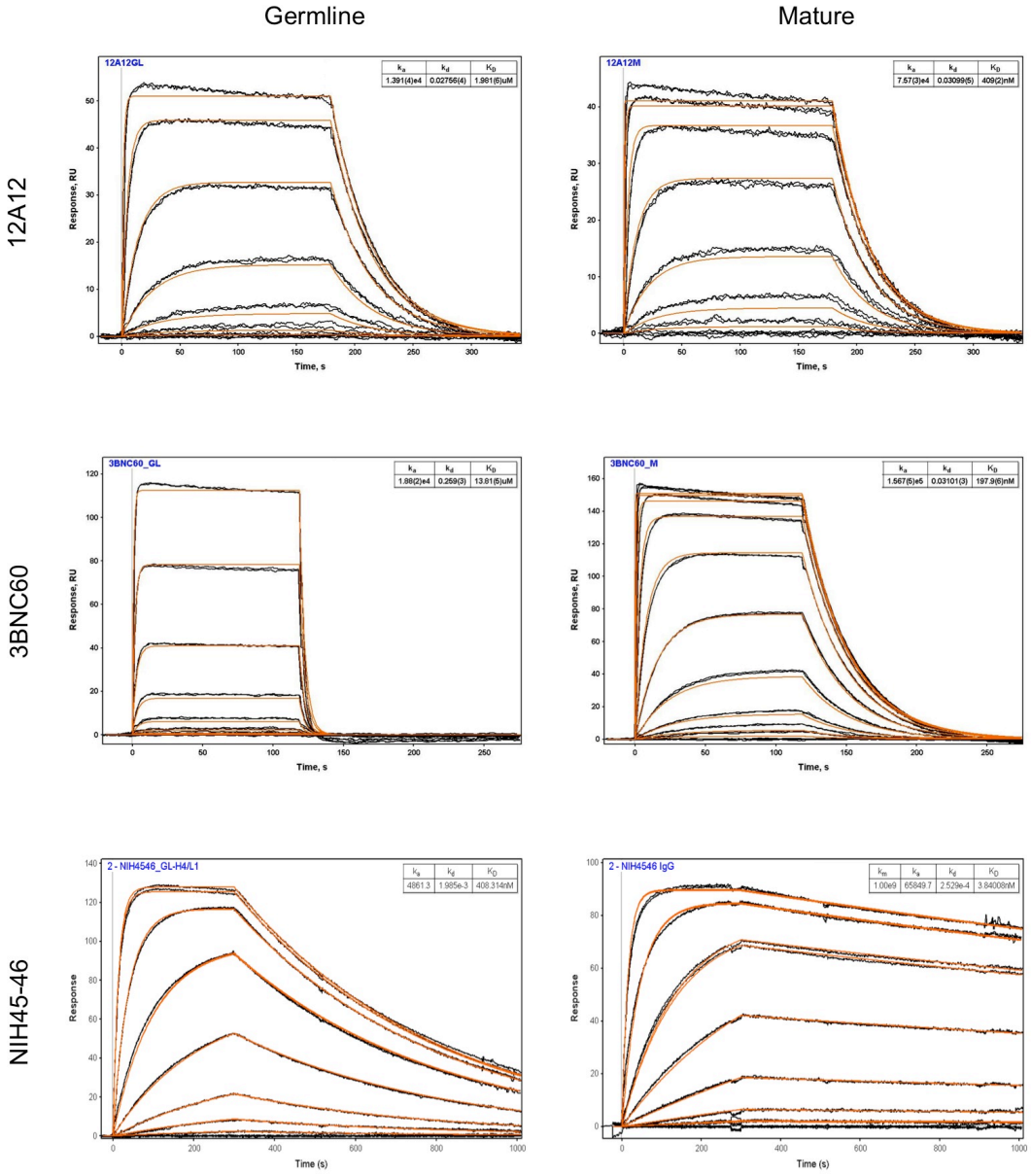


Fig. S25. Raw SPR data for eOD-GT6 against various GL and mature VRC01-class Abs binding to eOD-GT6.

Fig. S25. Cont.

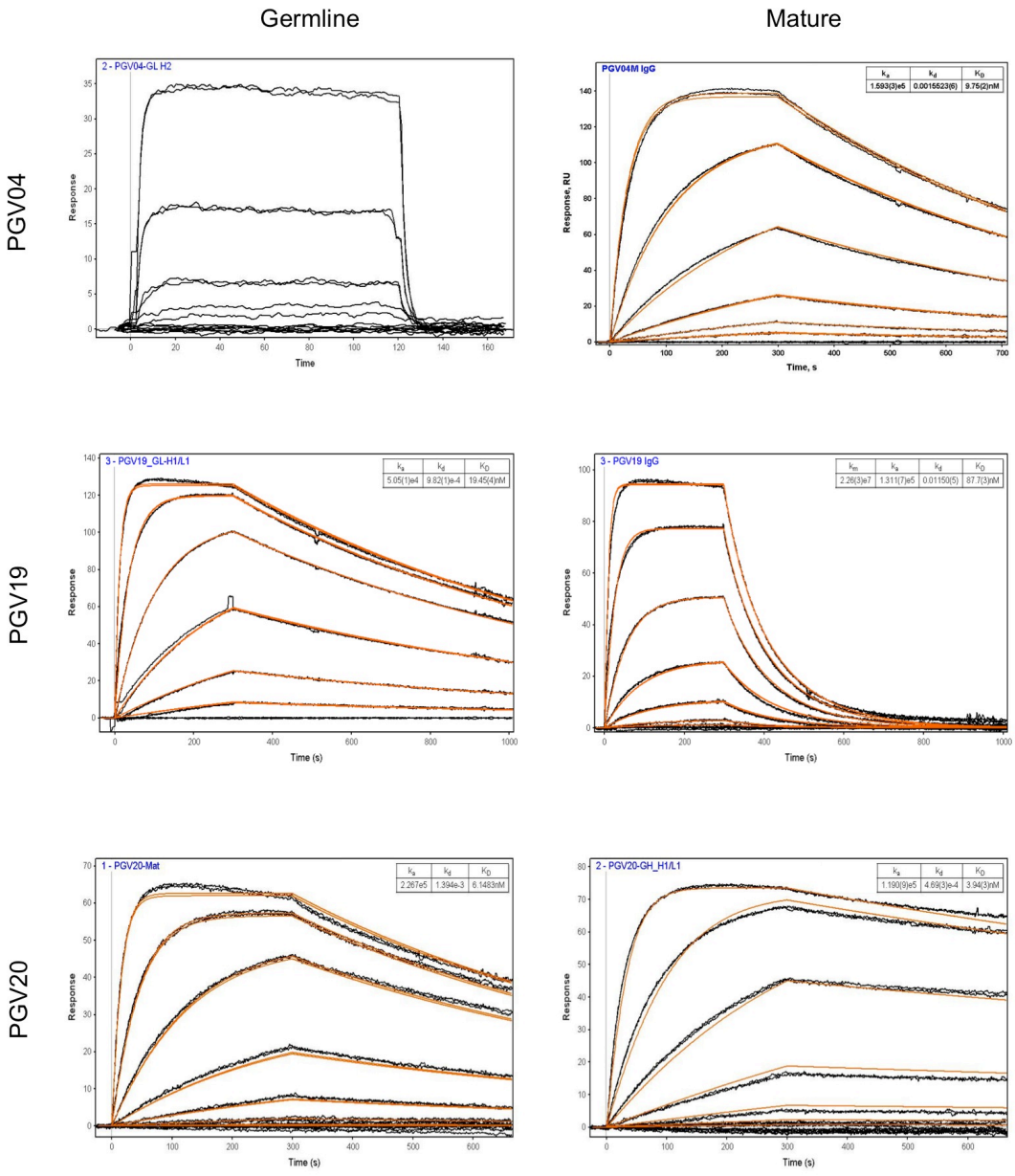
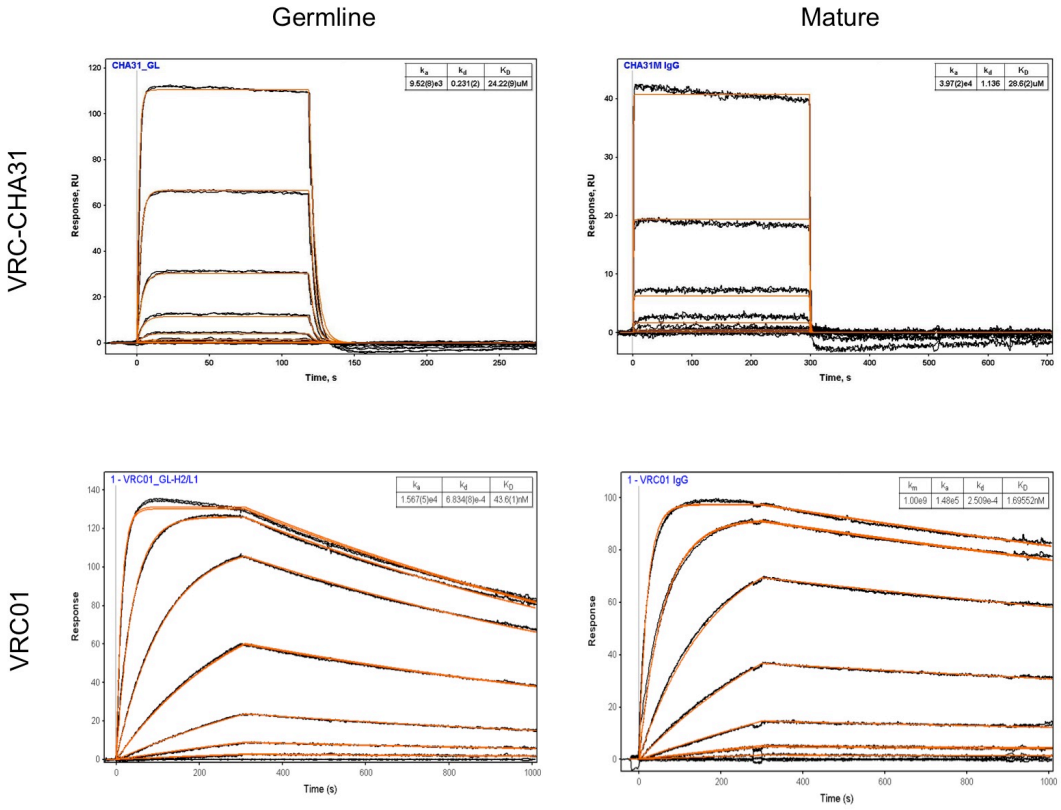

Germline
Mature
PGV04
PGV19
PGV20
2 - PGV04-GL H2
PGV04M IgG
3 - PGV19_GL-H1/L1
3 - PGV19 IgG
1 - PGV20-Mat
2 - PGV20-GH_H1/L1
Response
Response, RU
Time
Time, s
Time (s)

Fig. S25. Cont


Germline
Mature
VRC-CHA31
VRC01
CHA31_GL
CHA31M IgG
1 - VRC01_GL-H2/L1
1 - VRC01 IgG
Response, RU
Response
Time, s
Time (s)

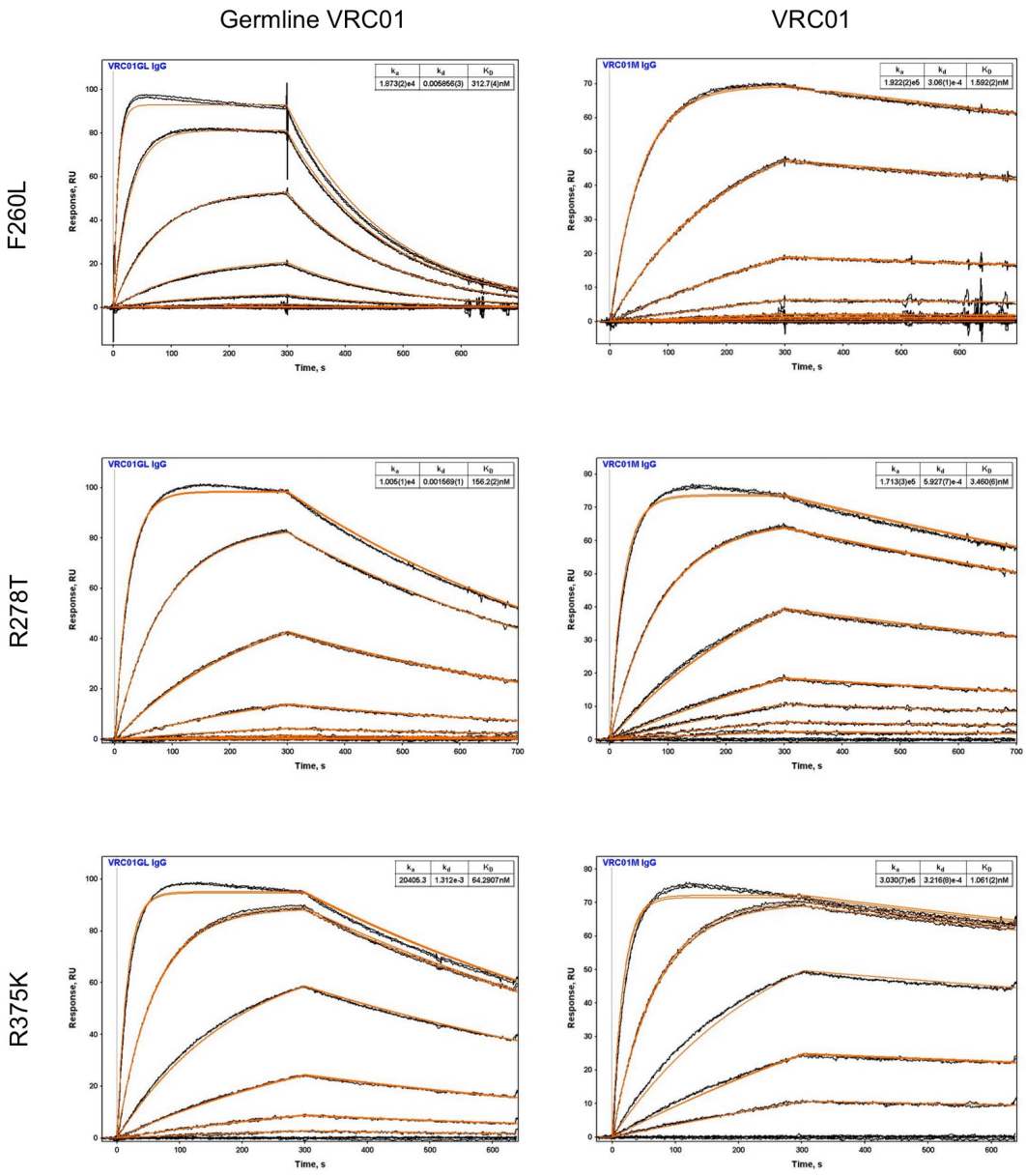
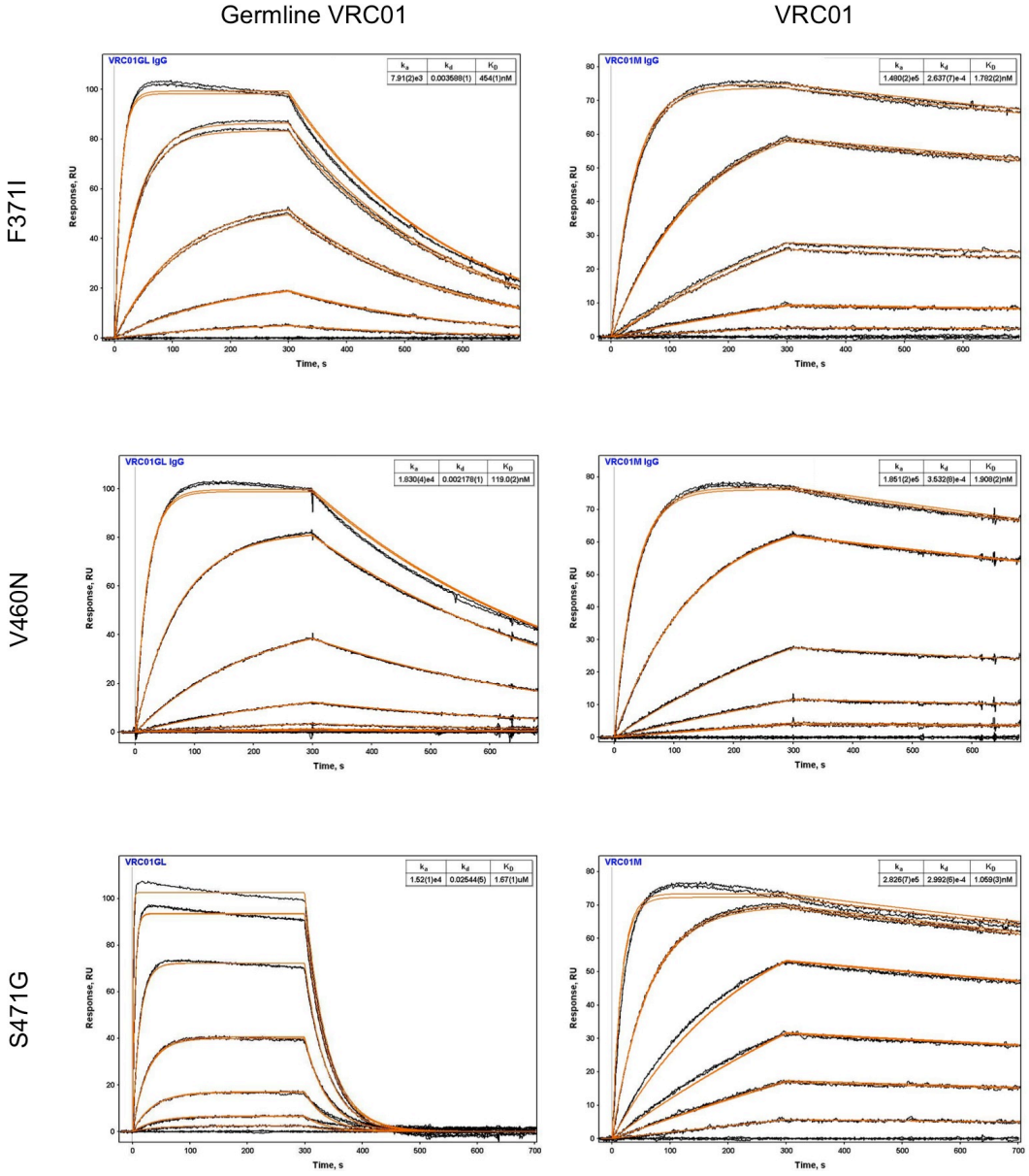


Fig. S26. Raw SPR data for GL and mature VRC01 binding to eOD-GT6 reversion mutants.

Fig. S26. Cont.


Germline VRC01
VRC01
F371I
V460N
S471G
VRC01GL IgG
VRC01M IgG
VRC01GL
VRC01M
Response, RU
Time, s

Supplementary Tables:

Table S1. Germline Binding ELISA

Clade	Virus Strain	VRC01	3BNC60	PGV19	GL VRC01	GL 3BNC60	GL PGV19
A	BL274.W6M.ENV.A3	1.31	0.23	0.78	>100	>100	>100
	Q23	2.17	0.38	0.85	>100	>100	>100
	Q769env.h5	1.37	0.29	0.27	>100	>100	>100
	92RW020	2.65	0.49	20.81	>100	>100	>100
	94UG103	3.46	1.82	1.30	>100	>100	>100
	QB726.70M.ENV.C4	0.55	0.40	>100	>100	>100	>100
	QF495.23M.ENV.A1	1.28	0.13	>100	>100	>100	>100
	QF495.23M.ENV.A3	1.11	0.11	>100	>100	>100	>100
	QF495.23M.ENV.B2	1.12	0.11	>100	>100	>100	>100
	QF495.23M.ENV.D1	0.77	0.06	>100	>100	>100	0.00
	QH209.14M.ENV.A2	0.81	1.83	>100	>100	>100	>100
	QH359.21M.ENV.D1	3.40	0.85	>100	>100	>100	>100
	QG984.21M.ENV.A3	0.75	>100	0.12	>100	>100	>100
	QH343.21M.ENV.A10	>100	>100	>100	>100	>100	>100
	BG505.W6M.ENV.C2	>100	>100	>100	>100	>100	>100
B	YU-2	1.02	0.09	0.44	>100	>100	>100
	6535.5	7.17	0.98	1.90	>100	>100	>100
	CAAN5342A2	5.24	0.94	3.07	>100	>100	>100
	TRO.11	3.39	0.56	7.76	>100	>100	>100
	TRJO4551.58	5.12	3.68	5.32	>100	>100	>100
	SC422661.8	1.66	0.79	1.33	>100	>100	>100
	QH0692	2.85	0.54	3.54	>100	>100	>100
	SF162	5.77	6.84	2.19	>100	>100	>100
	ADA	2.41	0.80	1.38	>100	>100	>100
	JR-FL	1.45	0.38	1.71	>100	>100	>100
	JR-CSF	1.20	0.13	1.45	>100	>100	>100
	PVO.4	3.03	0.31	>100	>100	>100	>100
	RHPA4259.7	0.94	0.16	>100	>100	>100	>100
	92BR020	2.63	0.97	>100	>100	>100	>100
	AC10	11.39	>100	>100	>100	>100	>100
	WITO4160.33	>100	>100	>100	>100	>100	>100
B (T/F)	62357_14_D3_4589	2.93	0.16	1.18	>100	>100	>100
BC	CNE53	1.14	0.40	>100	>100	>100	>100
	CNE20	>100	>100	>100	>100	>100	>100
C	ZM53M.PB12	1.42	0.22	0.28	>100	>100	>100
	IAVI C22	7.88	1.18	3.44	>100	>100	>100
	ZM109	3.21	0.50	>100	>100	>100	>100
	ZM197	1.52	4.40	>100	>100	>100	>100
	ZM214	2.15	0.22	>100	>100	>100	>100
	QB099.391M.ENV.B1	1.30	>100	>100	>100	>100	>100
	DU172	>100	>100	>100	>100	>100	>100
	DU422	>100	>100	>100	>100	>100	>100
	ZM233	>100	>100	>100	>100	>100	>100
C (T/F)	BF1266	0.62	0.12	0.07	>100	>100	>100
D	QA013.701.ENV.H1	5.40	22.61	3.98	>100	>100	>100
	QA013.70I.ENV.M12	0.88	0.33	>100	>100	>100	>100
G	X1193_C1	1.99	0.39	0.81	>100	>100	>100
	X1254_C3	1.05	0.18	2.94	>100	>100	>100
	X2131_C1_B5	6.02	3.02	4.67	>100	>100	>100
	P1981_C5_3	2.90	0.97	2.17	>100	>100	>100
	X2088_C9	>100	>100	>100	>100	>100	>100
A/D	MF535.W0M.ENV.D11	1.13	1.90	7.85	>100	>100	>100
	QA790.204I.ENV.A4	0.42	0.08	0.06	>100	>100	>100
	QA790.204I.ENV.C1	0.62	0.21	0.10	>100	>100	>100
	QA790.204I.ENV.C8	0.91	0.47	0.16	>100	>100	>100
A2/D	QG393.60M.ENV.A1	1.45	0.49	0.86	>100	>100	>100
	QG393.60M.ENV.B7	1.02	2.40	0.83	>100	>100	>100
AE	CRF01_AE clone 269	3.88	1.28	>100	>100	>100	>100
AG	263-8	2.17	0.35	1.33	>100	>100	>100
	T250-4	>100	>100	>100	>100	>100	>100
	235-47	>100	>100	>100	>100	>100	>100
C/D	BK184.W6M.ENV.D2	4.16	1.46	>100	>100	>100	>100

Table S2: Characteristics of germline-targeting eOD Generations

Characteristics of each generation of germline-targeting eOD. K_Ds were measured by SPR kinetic fits except where indicated by (*) in which case K_Ds were measured by fits to equilibrium association data. ND, not done.

		K_D (nM) measured by SPR for antibody-antigen interaction						
Antigen	Mutations from eOD Base	Antibody						
		VRC01		NIH45-46			PGV19	
		GL	Mat	GL	GL_v2	Mat	GL	Mat
eOD Base	0	$>10^5$	5	$>10^5$	54,000 *	14	$>10^5$	3,100
eOD-GT1	10	44,000 *	1	$>10^5$	ND	3	7,800*	1,000*
eOD-GT2	14	15,000 *	3	$>10^5$	ND	4	13,000*	6,600*
eOD-GT3	15	220	3	4,600	ND	4	390	1800
eOD-GT4	17	34	2	1,000	ND	3	27	1800
eOD-GT5	14	530	57	4,800	ND	110	300	19,000 *
eOD-GT6	8	44	2	410	9	4	19	88

Table S3: K_Ds for eOD-GT6 interactions with GL VRC01-class Abs with alternate human VH genes. Values in nM.

	eOD-GT6
VH1-2*01	$>10^5$
VH1-2*02	44
VH1-2*03	39
VH1-2*04	39
VH1-3*01	$>10^5$
VH1-8*01	32,000
VH1-46*01	$>10^5$

Table S4. X-Ray Crystallographic Table and Refinement Statistics

Crystal	GL-VRC01 Fab	eOD-GT6	eOD-GT6 + GL-VRC01 Fab
Data Collection	APS 23-ID	APS 23-ID	SSRL 12-2
Wavelength, Å	1.03320	1.03320	0.9795
Space group	H3	$P2_1$	C2
Unit cell a, b, c (Å)	172.44, 172.44, 92.43	44.96, 217.73, 44.99	179.26, 63.3, 61.3
α, β, γ (°)	90, 90, 120	90, 119.97, 90	90, 90.43, 90
Molecules per ASU	2	4	1 Fab + 1 eOD
Resolution (Å) *	40.0 – 2.1 (2.2 – 2.1)	40.0 – 2.5 (2.6 – 2.5)	40.0 – 2.4 (2.5 – 2.4)
Completeness*	100 (100)	99.9 (99.9)	93.6 (96.8)
Redundancy*	5.8 (5.8)	3.9 (3.9)	2.7 (3.0)
No. total reflections	347,661	99,546	74,140
No. unique reflections	59,809	25,751	25,368
I/σ*	13.6 (2.8)	8.1 (2.9)	9.4 (2.1)
R_{sym} †,*	8.6 (52.1)	14.2 (43.0)	8.5 (43.2)
Refinement statistics			
Resolution (Å)	40.0 – 2.1	40.0 – 2.5	40.0 – 2.4
No. reflections (total/R_{free})	58,400/2,914	24,371/1,336	25,310/1,253
R_{cryst}/R_{free} ‡, §	17.3/21.2	20.9/25.9	20.0/26.3
RMSD bond length (Å)	0.007	0.006	0.005
RMSD bond angles (°)	1.13	1.16	1.00
Protein atoms/ solvent	6,518/387	5,183/45	4,588/173

atoms			
Wilson B-value ($Å^2$)	30.0	44.9	57.3
Overall B-value ($Å^2$)	38.1	25.9	47.6
Average B-value protein ($Å^2$)	35.5	31.6	47.8
Average B-value solvent ($Å^2$)	59.5	20.4	42.2
Ramachandran			
Preferred %	98.1	90.3	93.8
Allowed %	1.7	8.3	5.4
PDB ID	XXX	XXX	XXX

* Values in parentheses are for the highest resolution shell.

† R_{sym}= Σ|I-<I>|/Σ<I>, where I is the observed intensity, and <I> is the average intensity of multiple observations of symmetry related reflections.

Table S5: Analysis of the buried surface area and H-bonds on VRC01/GL-VRC01 in the complexes of VRC01+gp120core and GL-VRC01+eOD-GT6.

Interfaces were calculated using PDBePISA (*29*).

	gp120 - VRC01		eOD-GT – GL-VRC01	
	Residue #	BSA (Å²)	Residue #	BSA (Å²)
FR1	E1	0	E1	41.6
	I2	0	I2	5.9
	V3	35.6	V3	0
CDR L1	Q27	23.1	Q27	26.6
	Y28 (H)	30.7	S28 (H)	13.4
	G29	0	V29	18.2
	A32	0	Y32	8.1
CDR L3	Y91 (H)	68.8	Y91 (H)	71.5
	E96 (H)	57.7	E96 (H)	57.8
	F97	55.7	F97	23.2
FR1	I30	15.6	T30	7.2
CDR H1	D31	0	G31	7.3
	T33	2.3	Y33	46.8
FR2	W47	0	W47	24.8
CDR H2	W50 (H)	38.1	W50 (H)	44.1
	K52 (H)	41.0	N52 (H)	28.3
	R53	57.5	N53	36.0
	G54 (H)	50.8	S54 (H)	74.0
	G55	16.5	G55	19.7
	A56	35.0	G56	30.0
	V57	35.9	T57 (H)	47.0
	N58 (H)	64.7	N58 (H)	66.3
	Y59 (H)	37.3	Y59	18.6
	A60	7.5	A60	14.8
	R61 (H)	160.8	Q61 (H)	146.7
	P62	20.5	K62	11.2
	Q64 (H)	43.1	Q64 (H)	52.2
FR3	R71 (H)	27.5	R71 (H)	24.6
	V73	26.1	T73	0
	Y74	52.7	S74	0
CDR H3	D99 (H)	45.6	D99 (H)	42.4
	Y100	19.3	Y100	16.8
	N100A	14.0	N100A	11.7
	W100B (H)	42.3	W100B (H)	38.8
Total BSA (Å²)		1125.7		1075.6

Table S6: Analysis of the buried surface area and H-bonds on gp120 core and eOD-GT6. Blue shaded areas show residues present in eOD-GT6.

	gp120core + VRC01		eOD-GT6 – GL-VRC01	
	Residue #	BSA (Å²)	Residue #	BSA (Å²)
	K97 (HS)	26.5		
β3	G124	38.8		
	G198	13.2		
Loop D	E275		V78	5.9
	N276	22.6	D79 (H)	31.4
	T278 (H)	123.0	R81 (H)	130.0
	N279 (H)	55.6	D82 (H)	54.1
	N280 (H)	68.4	N83 (H)	70.9
	A281 (H)	70.1	A84	66.6
	K282 (H)	30.9	K85 (H)	39.4
	T283	12.1	S86	0
	S365 (H)	62.2	S138 (H)	66.4
β15	G366	22.0	G139	17.7
	G367	24.0	G140	27.6
	D368 (H)	49.1	D141 (H)	54.1
α3	I371	44.2	F144	22.5
	W427	7.7		
β21	T430	57.8		
β23	T455	31.6	T25	24.5
	R456 (H)	5.7	R26 (H)	5.8
	D457 (H)	47.2	D27 (H)	47.4
	G458 (H)	49.8	G28 (H)	43.2
	G459 (H)	68.7	G29 (H)	61.4
	A460	37.8	V30	64.3
V5	N461	68.4	S31	17.5
	N462		N32 (H)	11.4
	T463	13.6	D33	0
	N465 (H)	9.7	T35 (H)	0.2
β24	E466	5.5	E36	2.9
	T467 (H)	15.4	I37	14.6
	R469 (H)	22.8	R39 (H)	24.2
	G472	8.3	G42 (H)	61.4
	G473	28.4	G43 (H)	22.1
	N474	16.8	D44 (H)	36.1
	K476	0.3	R46	47.4
α5	D477		D47	6.4
Total BSA (Å²)		1158.2		1102.0

Table S7: SPR data for all experiments.

ND, not determined. NB, no binding. SPR traces for eOD-GT6 bound to GL and Mature Abs as well as the 6 point mutations are shown in fig S24-25.

Construct	Antibody	*k*a	*k*d	K_D	Equilibrium
eOD-GT1	NIH45-46 GL	NB	NB	-	NB
eOD-GT1	NIH45-46 Mature	1.42E+05	4.24E-04	2.98E-09	ND
eOD-GT1	PGV19 GL	ND	ND	-	7.80E-06
eOD-GT1	PGV19 Mature	1.17E+05	1.26E-01	1.08E-06	1.04E-06
eOD-GT1	VRC01 GL	ND	ND	-	4.40E-05
eOD-GT1	VRC01 Mature	2.60E+05	3.67E-04	1.41E-09	ND
eOD-GT2	NIH45-46 GL	NB	NB	-	NB
eOD-GT2	NIH45-46 Mature	1.44E+05	5.34E-04	3.71E-09	ND
eOD-GT2	PGV19 GL	ND	ND	-	1.25E-05
eOD-GT2	PGV19 Mature	ND	ND	-	6.59E-06
eOD-GT2	VRC01 GL	ND	ND	-	1.50E-05
eOD-GT2	VRC01 Mature	2.21E+05	5.99E-04	2.71E-09	ND
eOD-GT3	NIH45-46 GL	5.33E+0	2.48E-02	4.65E-	ND

		3		06	
		1.18E+0		3.71E-	
eOD-GT3	NIH45-46 Mature	5	4.38E-04	09	ND
		4.50E+0		3.85E-	
eOD-GT3	PGV19 GL	4	1.73E-02	07	ND
		7.00E+0		1.83E-	
eOD-GT3	PGV19 Mature	4	1.28E-01	06	1.81E-06
		1.30E+0		2.21E-	
eOD-GT3	VRC01 GL	4	2.87E-03	07	ND
		2.00E+0		2.70E-	
eOD-GT3	VRC01 Mature	5	5.41E-04	09	ND
		3.43E+0		9.99E-	
eOD-GT4	NIH45-46 GL	3	3.42E-03	07	ND
		1.86E+0		2.79E-	
eOD-GT4	NIH45-46 Mature	5	5.18E-04	09	ND
		5.02E+0		2.77E-	
eOD-GT4	PGV19 GL	4	1.39E-03	08	ND
		4.86E+0		1.83E-	
eOD-GT4	PGV19 Mature	4	8.90E-02	06	1.79E-06
		1.39E+0		3.41E-	
eOD-GT4	VRC01 GL	4	4.73E-04	08	ND
		2.59E+0		2.15E-	
eOD-GT4	VRC01 Mature	5	5.55E-04	09	ND
eOD-GT5	NIH45-46 GL	5.81E+0	2.79E-03	4.81E-	ND

		2		06	
		1.05E+0		1.11E-	
eOD-GT5	NIH45-46 Mature	4	1.16E-03	07	ND
		5.20E+0		3.01E-	
eOD-GT5	PGV19 GL	3	1.56E-03	07	ND
		3.90E+0		1.95E-	
eOD-GT5	PGV19 Mature	3	7.60E-02	05	1.92E-05
		1.51E+0		5.29E-	
eOD-GT5	VRC01 GL	3	7.99E-04	07	ND
		2.39E+0		5.70E-	
eOD-GT5	VRC01 Mature	4	1.36E-03	08	ND
		1.39E+0		1.98E-	
eOD-GT6	12A12 GL	4	2.76E-02	06	1.86E-06
		7.57E+0		4.09E-	
eOD-GT6	12A12 Mature	4	3.10E-02	07	3.71E-07
		1.88E+0		1.38E-	
eOD-GT6	3BNC60 GL	4	2.59E-01	05	1.36E-05
		1.57E+0		1.98E-	
eOD-GT6	3BNC60 Mature	5	3.10E-02	07	1.85E-07
		4.86E+0		4.08E-	
eOD-GT6	NIH45-46 GL	3	1.99E-03	07	ND
		6.58E+0		3.84E-	
eOD-GT6	NIH45-46 Mature	4	2.53E-04	09	ND
eOD-GT6	PGV04 GL	ND	ND	-	5.20E-05
eOD-GT6	PGV04 Mature	1.59E+0	1.55E-03	9.73E-	ND

		5		09	
		5.05E+0		1.94E-	
eOD-GT6	PGV19 GL	4	9.82E-04	08	ND
		1.31E+0		8.78E-	
eOD-GT6	PGV19 Mature	5	1.15E-02	08	ND
		1.19E+0		3.94E-	
eOD-GT6	PGV20 GL	5	4.69E-04	09	ND
		2.27E+0		6.15E-	
eOD-GT6	PGV20 Mature	5	1.39E-03	09	ND
		9.52E+0		2.43E-	
eOD-GT6	VRC-CH31 GL	3	2.31E-01	05	2.38E-05
	VRC-CH31	3.97E+0	1.14E+0	2.86E-	
eOD-GT6	Mature	4	0	05	2.84E-05
		1.57E+0		4.36E-	
eOD-GT6	VRC01 GL	4	6.83E-04	08	ND
		1.48E+0		1.69E-	
eOD-GT6	VRC01 Mature	5	2.51E-04	09	ND
eOD-GT6	VRC03 GL	NB	NB	-	NB
		1.79E+0		8.92E-	
eOD-GT6	VRC03 Mature	5	1.59E-01	07	8.72E-07
eOD-GT6	VRC06 GL	NB	NB	-	NB
		5.53E+0		2.19E-	
eOD-GT6	VRC06 Mature	4	1.21E-01	06	2.10E-06
		1.83E+0		1.19E-	
eOD-GT6 V460N	VRC01 GL	4	2.18E-03	07	ND

eOD-GT6 V460N	VRC01 Mature	1.85E+05	3.53E-04	1.91E-09	ND
eOD-GT6 S471G	VRC01 GL	1.54E+04	2.54E-02	1.65E-06	1.37E-06
eOD-GT6 S471G	VRC01 Mature	2.83E+05	2.99E-04	1.06E-09	ND
eOD-GT6 F260L	VRC01 GL	1.87E+04	5.86E-03	3.13E-07	ND
eOD-GT6 F260L	VRC01 Mature	1.92E+05	3.06E-04	1.59E-09	ND
eOD-GT6 R278T	VRC01 GL	1.00E+04	1.57E-03	1.57E-07	ND
eOD-GT6 R278T	VRC01 Mature	1.71E+05	5.93E-04	3.46E-09	ND
eOD-GT6 R357K	VRC01 GL	2.04E+04	1.31E-03	6.43E-08	ND
eOD-GT6 R357K	VRC01 Mature	3.03E+05	3.22E-04	1.06E-09	ND
eOD-GT6 F371I	VRC01 GL	7.91E+03	3.59E-03	4.54E-07	ND
eOD-GT6 F371I	VRC01 Mature	1.48E+05	2.64E-04	1.78E-09	ND
eOD-GT6 T465S	VRC01 GL	1.13E+03	4.65E-04	4.13E-08	ND
eOD-GT6 T465S	VRC01 Mature	9.90E+04	2.46E-04	2.50E-09	ND

eOD-GT6	D368R		NIH45-46 GL	NB	NB	-	NB
eOD-GT6	D368R		NIH45-46 Mature	NB	NB	-	NB
eOD-GT6	D368R		PGV19 GL	1.39E+04	1.93E-01	1.39E-05	ND
eOD-GT6	D368R		PGV19 Mature	ND	ND	-	1.93E-05
eOD-GT6	D368R		VRC01 GL	2.92E+03	8.59E-02	2.94E-05	ND
eOD-GT6	D368R		VRC01 Mature	ND	ND	-	4.00E-05
eOD-GT6	D279A		NIH45-46 GL	NB	NB	-	NB
eOD-GT6	D279A		NIH45-46 Mature	NB	NB	-	NB
eOD-GT6	D279A		PGV19 GL	NB	NB	-	NB
eOD-GT6	D279A		PGV19 Mature	NB	NB	-	NB
eOD-GT6	D279A		VRC01 GL	NB	NB	-	NB
eOD-GT6	D279A		VRC01 Mature	ND	ND	-	1.20E-04
eOD-GT6 D463N	D386N	+	VRC01 GL	6.47E+03	7.99E-04	1.24E-07	ND
eOD-GT6 D463N	D386N	+	VRC01 Mature	5.55E+04	4.44E-04	7.99E-09	ND
eOD-GT6 D463N	D386N	+	NIH45-46 GL	2.20E+03	2.38E-03	1.08E-06	ND
eOD-GT6 D463N	D386N	+	NIH45-46 Mature	2.19E+04	7.71E-04	3.52E-08	ND
eOD-GT6 D463N	D386N	+	PGV19 GL	1.93E+04	1.07E-03	5.54E-08	ND
eOD-GT6	D386N	+	PGV19 Mature	2.00E+03	2.00E+03	1.00E-	ND

D463N		6	0	06	
eOD-GT6	VRC01_VH1-2*01	NB	NB	-	NB
eOD-GT6	VRC01_VH1-2*03	1.59E+04	6.13E-04	3.86E-08	ND
eOD-GT6	VRC01_VH1-2*04	1.50E+04	5.77E-04	3.85E-08	ND
eOD-GT6	VRC01_VH1-3*01	NB	NB	-	NB
eOD-GT6	VRC01_VH1-8*01	ND	ND	-	3.20E-04
eOD-GT6	VRC01_VH1-46*01	NB	NB	-	NB
eOD-GT6	VRC01_Rhe1	NB	NB	-	NB
eOD-GT6	VRC01_Rhe2	ND	ND	-	2.74E-05
eOD-GT6	VRC01_Rhe4	ND	ND	-	3.78E-05
eOD Base	12A12 GL	NB	NB	-	NB
eOD Base	12A12 Mature	5.28E+04	2.00E-02	3.79E-07	ND
eOD Base	3BNC60 GL	NB	NB	-	NB
eOD Base	3BNC60 Mature	-	-	-	-
eOD Base	NIH45-46 GL	NB	NB	-	NB
eOD Base	NIH45-46 Mature	4.34E+04	6.18E-04	1.43E-08	ND
eOD Base	PGV04 GL	NB	NB	-	NB
eOD Base	PGV04 Mature	4.57E+04	5.18E-03	1.13E-07	ND

eOD Base	PGV19 GL	NB	NB	-	NB
eOD Base	PGV19 Mature	5.07E+04	1.58E-01	3.12E-06	2.99E-06
eOD Base	PGV20 GL	ND	ND	-	1.70E-05
eOD Base	PGV20 Mature	1.89E+05	2.54E-03	1.34E-08	
eOD Base	VRC-CH31 GL	NB	NB	-	NB
eOD Base	VRC-CH31 Mature	ND	ND	-	3.02E-05
eOD Base	VRC01 GL	NB	NB	-	NB
eOD Base	VRC01 Mature	7.05E+04	3.51E-04	4.98E-09	ND
HXB2 Core GP120	12A12 GL	NB	NB	-	NB
HXB2 Core GP120	12A12 Mature	2.60E+04	1.67E-04	6.42E-09	ND
HXB2 Core GP120	3BNC60 GL	NB	NB	-	NB
HXB2 Core GP120	3BNC60 Mature	1.17E+04	4.22E-04	3.62E-08	ND
HXB2 Core GP120	NIH45-46 GL	NB	NB	-	NB
HXB2 Core GP120	NIH45-46 Mature	1.30E+04	4.56E-04	3.50E-08	ND
HXB2 Core GP120	PGV04 GL	NB	NB	-	NB
HXB2 Core GP120	PGV04 Mature	8.86E+03	4.28E-04	4.83E-08	ND
HXB2 Core GP120	PGV19 GL	NB	NB	-	NB

		5.62E+0		2.02E-	
HXB2 Core GP120	PGV19 Mature	3	1.13E-04	08	ND
HXB2 Core GP120	PGV20 GL	NB	NB	-	NB
		8.97E+0		1.90E-	
HXB2 Core GP120	PGV20 Mature	4	1.71E-03	08	ND
HXB2 Core GP120	VRC-CH31 GL	NB	NB	-	NB
	VRC-CH31	4.23E+0		4.71E-	
HXB2 Core GP120	Mature	3	1.99E-04	08	ND
HXB2 Core GP120	VRC01 GL	NB	NB	-	ND
		1.71E+0		4.54E-	
HXB2 Core GP120	VRC01 Mature	4	7.74E-05	09	ND
		2.61E+0		3.17E-	
Core.Bal-GT1	12A12 GL	4	8.27E-02	06	3.14E-06
		1.56E+0		1.04E-	
Core.Bal-GT1	12A12 Mature	5	1.62E-04	09	ND
		4.65E+0		1.64E-	
Core.Bal-GT1	3BNC60 GL	4	7.64E-01	05	1.59E-05
		2.57E+0		3.94E-	
Core.Bal-GT1	3BNC60 Mature	5	1.01E-03	09	ND
Core.Bal-GT1	NIH45-46 GL	NB	NB	-	NB
		1.10E+0		6.37E-	
Core.Bal-GT1	NIH45-46 Mature	5	7.01E-05	10	ND
		1.18E+0		2.55E-	
Core.Bal-GT1	PGV19 GL	5	3.00E-03	08	ND
Core.Bal-GT1	PGV19 Mature	7.40E+0	1.01E-02	1.36E-	ND

		5		08	
		7.58E+0		5.55E-	
Core.Bal-GT1	PGV20 GL	3	4.21E-02	06	ND
		2.66E+0		3.73E-	
Core.Bal-GT1	PGV20 Mature	5	9.93E-04	09	ND
		1.36E+0		3.54E-	
Core.Bal-GT1	VRC-CH31 GL	4	4.81E-01	05	3.30E-05
	VRC-CH31	3.12E+0		1.79E-	
Core.Bal-GT1	Mature	4	5.58E-02	06	2.00E-06
		3.00E+0		1.33E-	
Core.Bal-GT1	VRC01 GL	5	4.00E-01	06	1.80E-06
		1.96E+0		4.93E-	
Core.Bal-GT1	VRC01 Mature	5	9.66E-05	10	ND

Table S8: K_Ds for eOD-GT6 interactions with GL VRC01-class Abs with selected non-human VH genes. Values in nM.

	eOD-GT6
Mus VH4-1	$>10^5$
Mus VH5-4	$>10^5$
Mus VH5-12	$>10^5$
Mus VH5-17	$>10^5$
Mus VH11-1	$>10^5$
Rhe1	$>10^5$
Rhe2	30,000
Rhe4	40,000

Table S9. Strains in LANL Lacking the N276 Glycan

HIV strains lacking glycan at 276 out of 2867 different sequences from the LANL database:

Strain_name	276	277	278
A.CM.97.97CM_MP640.AM279366	S	L	T
A.UG.07.191955_A11.HM215272	N	I	N
A1.KE.00.KNH1211.AF457070	N	I	A
A1.KE.04.QG984_21M_ENV_A3.FJ866117	D	I	S
A1.TZ.01.A341.AY253314	N	I	A
B.AU.96.MBCD36.AF042105	N	F	M
B.CH.x.NAB8pre_cl_11.EU023927	D	F	S
B.CN.06.CNHLJSM06048.EU131794	N	F	M
B.CN.x.B05.EU363829	N	F	M
B.CO.01.PCM039.AY561239	N	F	A
B.GB.04.MM39d11p.HM586193	N	F	M
B.GB.08.F455b_B4.HQ595778	N	V	E
B.GB.x.MB314.Y13719	N	F	M
B.JP.05.-426.AB428556	N	F	X
B.US.01.108051_006.HM769944	D	F	K
B.US.06.YOMI_A6.EU578667	N	F	A
B.US.87.SFMHS5.AF025753	D	F	T
B.US.97.1013_03.AY331287	N	F	X
B.US.97.62357_14_D3_4589.EU289189	K	F	A
B.US.97.M02_3_SW.U84854	D	F	S
B.US.x.F1540TOB8U.GU728134	S	F	T
B.US.x.H0002GH.DQ222211	D	F	K
B.US.x.L3516TOB8U.GU728392	D	F	M
C.BW.00.00BW076820.AF443089	A	L	T
C.BW.00.00BW087421.AF443090	N	L	A
C.BW.00.00BW18595.AF443099	N	I	E
C.BW.00.00BW20361.AF443102	N	L	A
C.BW.96.96BW11B01.AF110971	I	I	T
C.BW.99.99BW47547.AF443086	S	L	T
C.GE.03.03GEMZ033.DQ207941	D	I	N
C.IN.00.HIV_00836_2.EF117265	K	L	D
C.IN.x.CALCMANDAL.AJ276221	N	L	I

C.MW.03.CHV0011247_0478_H2.FJ444251	E	L	G
C.MW.x.BF1266_431a.HM215360	N	L	E
C.TZ.00.390_F1_B7.HQ697983	N	L	K
C.TZ.03.6471_v1_c16.HM215328	D	L	N
C.TZ.x.346_F4_D2_12.HM215302	N	I	N
C.ZA.00.1184MB.AY838566	N	L	A
C.ZA.00.1189MB.AY838565	N	L	A
C.ZA.00.1225MB.AY463227	N	L	I
C.ZA.00.J112MA.AY838568	N	L	K
C.ZA.02.02ZAPS005MB1.DQ351235	D	L	T
C.ZA.02.02ZAPS013MB1.DQ351222	N	L	A
C.ZA.02.02ZAPS014MB1.DQ351218	N	L	A
C.ZA.03.03ZAPS032MB1.DQ445633	N	L	A
C.ZA.03.03ZAPS056MB1.DQ396374	N	L	K
C.ZA.03.03ZAPS066MB2.DQ396375	N	M	K
C.ZA.03.03ZAPS067MB2.DQ396389	N	L	E
C.ZA.03.03ZAPS116MB1.DQ445635	N	L	A
C.ZA.03.03ZAPS136MB1.DQ351231	N	L	A
C.ZA.03.03ZAPS151MB1.DQ396392	K	L	T
C.ZA.03.03ZASK034B1.AY878065	N	P	T
C.ZA.03.03ZASK078B1.AY901971	K	L	T
C.ZA.03.03ZASK098B1.AY878061	N	L	A
C.ZA.03.03ZASK224MB1.DQ275664	K	L	T
C.ZA.03.03ZASK232B1.DQ093589	D	L	T
C.ZA.04.04ZAPS157MB1.DQ351232	N	L	A
C.ZA.04.04ZASK139B1.AY878072	D	L	T
C.ZA.04.04ZASK168B1.AY878058	D	L	T
C.ZA.04.04ZASK181B1.AY878062	S	L	T
C.ZA.04.04ZASK204B1.DQ056414	N	L	A
C.ZA.04.SK144B1.AY703911	N	L	A
C.ZA.05.05ZAFV5.DQ382363	N	L	A
C.ZA.05.CAP174_4w.GQ999981	N	L	A
C.ZA.05.CAP239_5w_F1.GQ999991	N	I	L
C.ZA.06.2833264.HQ595757	N	L	A
C.ZA.06.CHV0005989_CAP129.1.15B2.FJ443417	K	L	E
C.ZA.06.CHV0006091_CAP222.1.11A6.FJ443515	D	L	E

C.ZA.07.3514597.HQ595759	N	M	A
C.ZA.07.CHV0008598_CAP40.2.01F2.FJ443865	N	L	K
C.ZA.09.704MC004N.GU080162	S	L	T
C.ZA.09.704MC016N.GU080173	D	L	E
C.ZA.98.TV013B.AF391246	S	L	E
C.ZA.99.COT9.DQ447272	N	I	K
C.ZA.99.LT36.AY522729	N	L	A
C.ZM.02.14M_BML_1012.HM036760	N	L	A
C.ZM.02.21M_BML_1012.GU939143	N	L	A
C.ZM.04.Z205FPB5NOV04ENV5.2.GQ485436	N	L	A
D.TZ.87.87TZ4622.U65075	S	L	T
D.UG.97.9FPC2.EU853126	N	L	A
D.UG.97.pt632.EU281996	S	L	T
D.UG.98.98UG57131.AF484505	N	I	K
D.UG.99.99UGA07412.AF484477	N	X	X
D.UG.99.99UGD23550.AF484485	K	L	E
D.YE.01.01YE386.AY795903	N	L	A
F1.RU.08.D88_845.GQ290462	N	I	K
K.CD.97.97ZR_EQTB11.AJ249235	D	I	T
U.NL.01.U_NL_01_H10986_C11.EF029069	N	I	I
01_AE.CF.90.90CF402.U51188	D	L	T
01_AE.CN.06.CNE3.HM215410	D	L	T
01_AE.CN.07.BJX4_6.GU475020	N	L	A
01_AE.TH.00.00TH_C3347.AY945721	N	L	E
01_AE.TH.01.OUR414I.AY358050	N	L	A
01_AE.TH.06.99PL2.EU743787	N	I	K
01_AE.TH.95.NI1144.AF070703	N	L	A
01_AE.TH.97.97TH6_107.AY125894	N	P	T
02_AG.CM.01.01CM_0119MA.GU201498	D	I	T
02_AG.CM.01.01CM_0186BA.GU201500	D	I	T
02_AG.CM.01.01CM_4410HAL.AY371142	S	I	T
02_AG.CM.02.02CM_4082STN.AY371141	S	I	T
02_AG.CM.04.250.EU513189	N	T	I
02_AG.US.99.99US_MSC1134.AY444809	S	L	T
05_DF.BE.93.VI961.AF076998	N	I	L
07_BC.CN.x.CH110.EF117257	K	L	T

14_BG.ES.00.X477.AF423759	N	F	X
14_BG.ES.05.X772_8.FJ670528	N	F	X
17_BF.PE.02.PE02_PCR0155.EU581828	N	I	F
24_BG.ES.08.X2456_2.FJ670526	N	F	A
35_AD.AF.06.047H.GQ477443	N	I	L
01B.MY.04.04MYKL016_1.DQ366663	X	L	T
02A1.GH.97.AG2.AB052867	N	L	A
02A1U.CM.05.280.EU513183	D	I	T
02G.CM.02.02CM_3228MN.AY371147	N	I	L
A1C.TZ.97.97TZ06.AF361876	D	L	T
A1D.DK.96.FSA.DQ912822	D	I	L
A1D.UG.98.98UG57129.AF484503	N	I	I
AC.TZ.04.6540_v4_c1.HM215330	H	I	G
AC.TZ.x.605_F4_12_5.HM215322	D	L	E
BF1.BR.99.BREPM107.AY771588	N	I	K
BF1.ES.02.X1241.AY536238	N	I	A
BF1.IT.05.83166.GU595152	N	I	X
CD.TZ.x.89_F1_2_25.HM215349	E	I	T
CF1.MW.07.CH010180_w12_p1.HM204593	N	L	K
DF.CM.93.CA4.AJ277819	N	I	R
DU.CM.x.247.EU683891	D	F	T
GKU.SE.95.SE9010.AY352655	D	M	T
O.BE.87.ANT70.L20587	D	I	L
O.CM.96.96ABB009.AF383231	D	I	K
O.CM.98.98CMA124.AF383245	N	I	M
O.CM.98.98CMA307.AF383246	D	I	R
O.CM.98.98CMA323.AF383248	S	I	T
O.CM.98.98CMA407.AF383250	N	I	F
O.FR.92.VAU.AF407418	D	I	S
N.CM.02.DJO0131.AY532635	-	-	N
N.CM.02.SJGddd.GQ324959	-	-	N
N.CM.04.04CM_1015_04.DQ017382	-	-	X
N.CM.06.U14296.GQ324962	-	-	-
N.CM.06.U14842.GQ324958	S	D	S
N.CM.95.YBF30.AJ006022	-	-	N
N.CM.97.YBF106.AJ271370	-	-	-

CPZ.CD.90.ANT.U42720	R	K	N
CPZ.CM.05.SIVcpzEK505.DQ373065	-	-	N
CPZ.CM.05.SIVcpzLB7.DQ373064	-	-	-
CPZ.CM.98.CAM3.AF115393	-	-	-
CPZ.CM.98.CAM5.AJ271369	D	L	R

References:

1. S. A. Plotkin, Correlates of protection induced by vaccination. *Clin Vaccine Immunol* **17**, 1055 (2010).
2. D. R. Burton, P. Poignard, R. L. Stanfield, I. A. Wilson, Broadly neutralizing antibodies present new prospects to counter highly antigenically diverse viruses. *Science* **337**, 183 (2012).
3. W. R. Schief, Y.-E. A. Ban, L. Stamatatos, Challenges for structure-based HIV vaccine design. *Curr Opin HIV AIDS* **4**, 431 (2009).
4. D. R. Burton *et al.*, A blueprint for HIV vaccine discovery. *Cell Host Microbe* **12**, 396 (2012).
5. D. C. Ekiert *et al.*, Antibody recognition of a highly conserved influenza virus epitope. *Science* **324**, 246 (2009).
6. P. Zhang *et al.*, Depletion of interfering antibodies in chronic hepatitis C patients and vaccinated chimpanzees reveals broad cross-genotype neutralizing activity. *Proc Natl Acad Sci U S A* **106**, 7537 (2009).
7. J. P. Julien, P. S. Lee, I. A. Wilson, Structural insights into key sites of vulnerability on HIV-1 Env and influenza HA. *Immunol Rev* **250**, 180 (2012).
8. L. Kong *et al.*, Structural basis of hepatitis C virus neutralization by broadly neutralizing antibody HCV1. *Proc Natl Acad Sci U S A* **109**, 9499 (2012).
9. D. R. Burton, Antibodies, viruses and vaccines. *Nat Rev Immunol* **2**, 706 (2002).
10. X. L. Wu *et al.*, Rational design of envelope identifies broadly neutralizing human monoclonal antibodies to HIV-1. *Science* **329**, 856 (2010).
11. J. F. Scheid *et al.*, Sequence and structural convergence of broad and potent HIV antibodies that mimic CD4 binding. *Science* **333**, 1633 (2011).
12. X. Wu *et al.*, Focused evolution of HIV-1 neutralizing antibodies revealed by structures and deep sequencing. *Science* **333**, 1593 (2011).
13. T. Q. Zhou *et al.*, Structural basis for broad and potent neutralization of HIV-1 by antibody VRC01. *Science* **329**, 811 (2010).
14. R. Diskin *et al.*, Increasing the potency and breadth of an HIV antibody by using structure-based rational design. *Science* **334**, 1289 (2011).
15. A. P. West, Jr., R. Diskin, M. C. Nussenzweig, P. J. Bjorkman, Structural basis for germ-line gene usage of a potent class of antibodies targeting the CD4-binding site of HIV-1 gp120. *Proc Natl Acad Sci U S A* **109**, E2083 (2012).
16. R. Arnaout *et al.*, High-resolution description of antibody heavy-chain repertoires in humans. *Plos One* **6**, (2011).
17. R. Pejchal *et al.*, A potent and broad neutralizing antibody recognizes and penetrates the HIV glycan shield. *Science* **334**, 1097 (2011).
18. A. Leaver-Fay *et al.*, ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol* **487**, 545 (2011).
19. G. Chao *et al.*, Isolating and engineering human antibodies using yeast surface display. *Nature Protocols* **1**, 755 (2006).
20. C. Genomes Project *et al.*, An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56 (2012).
21. M. F. Bachmann, G. T. Jennings, Vaccine delivery: a matter of size, geometry, kinetics and molecular patterns. *Nat Rev Immunol* **10**, 787 (2010).
22. X. Zhang, W. Meining, M. Fischer, A. Bacher, R. Ladenstein, X-ray structure analysis and crystallographic refinement of lumazine synthase from the hyperthermophile Aquifex aeolicus at 1.6 A resolution: determinants of thermostability revealed from structural comparisons. *J Mol Biol* **306**, 1099 (2001).
23. T. Ota *et al.*, Anti-HIV B cell lines as candidate vaccine biosensors. *J Immunol*, (2012).
24. S. Hoot *et al.*, Recombinant HIV Envelope Proteins Fail to Engage Germline Versions of Anti-CD4bs bNAbs. *PLoS Pathog* **9**, e1003106 (2013).

25. C. Sundling *et al.*, Soluble HIV-1 Env trimers in adjuvant elicit potent and diverse functional B cell responses in primates. *J Exp Med* **207**, 2003 (2010).
26. K. L. Knight, C. R. Winstead, Generation of antibody diversity in rabbits. *Curr Opin Immunol* **9**, 228 (1997).
27. Several mouse VH genes were identified that contained the ArgH71 important for eOD-GT6 binding, but crystal structures of mouse Abs indicated that ArgH71 would not be accessible for binding, as it forms internal H-bonds with backbone carbonyl groups (fig. S15).
28. B. F. Haynes, G. Kelsoe, S. C. Harrison, T. B. Kepler, B-cell-lineage immunogen design in vaccine development with HIV-1 as a case study. *Nat Biotechnol* **30**, 423 (2012).
29. E. Krissinel, K. Henrick, Inference of macromolecular assemblies from crystalline state. *J Mol Biol* **372**, 774 (2007).
30. E. F. Pettersen *et al.*, UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605 (2004).
31. M. M. Souto-Carneiro, N. S. Longo, D. E. Russ, H. W. Sun, P. E. Lipsky, Characterization of the human Ig heavy chain antigen binding complementarity determining region 3 using a newly developed software algorithm, JOINSOLVER. *J Immunol* **172**, 6790 (2004).
32. P.-S. Huang *et al.*, RosettaRemodel: a generalized framework for flexible backbone protein design. *Plos One* **6**, (2011).
33. B. Kuhlman *et al.*, Design of a novel globular protein fold with atomic-level accuracy. *Science* **302**, 1364 (2003).
34. S. J. Fleishman *et al.*, RosettaScripts: a scripting language interface to the rosetta macromolecular modeling suite. *Plos One* **6**, (2011).
35. J. J. Gray *et al.*, Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. *J Mol Biol* **331**, 281 (2003).
36. I. W. Davis, W. B. Arendall, D. C. Richardson, J. S. Richardson, The backrub motion: How protein backbone shrugs when a sidechain dances. *Structure* **14**, 265 (2006).
37. W. P. Stemmer, A. Crameri, K. D. Ha, T. M. Brennan, H. L. Heyneker, Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. *Gene* **164**, 49 (1995).
38. L. Benatuil, J. M. Perez, J. Belk, C.-M. Hsieh, An improved yeast transformation method for the generation of very large human antibody libraries. *Protein Eng Des Sel* **23**, 155 (2010).
39. S. Bryson *et al.*, Cross-neutralizing human monoclonal anti-HIV-1 antibody 2F5: Preparation and crystallographic analysis of the free and epitope-complexed forms of its F-ab ' fragment. *Protein Peptide Lett* **8**, 413 (2001).
40. R. S. Depetris *et al.*, Partial enzymatic deglycosylation preserves the structure of cleaved recombinant HIV-1 envelope glycoprotein trimers. *J Biol Chem* **287**, 24239 (2012).
41. W. Kabsch, Xds. *Acta Crystallogr D Biol Crystallogr* **66**, 125 (2010).
42. A. J. McCoy *et al.*, Phaser crystallographic software. *J Appl Crystallogr* **40**, 658 (2007).
43. M. D. Winn *et al.*, Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **67**, 235 (2011).
44. P. D. Adams *et al.*, PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213 (2010).
45. P. Emsley, K. Cowtan, Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126 (2004).

CHAPTER 5: WILDTYPE EOD MULTIMERIZATION

In the previous chapter, we have shown that multimerizaton of a germline targeted eOD construct is extremely beneficial, allowing eOD-GT6 60mer to activate germline VCR01 B cells at a concentration as low as 1 nM. While the germline-targeted construct was designed to bind both mature and germline versions of VRC01, the sequence on the surface of the eOD-GT6 60mer differs from the native HIV gp120. If the antibodies produced will be specific to the engineered sequence and fail to recognize the native sequence and its respective conformation, they will not be able to broadly cross-react to the various HIV strains. Hence, the eOD-GT6 construct may be a fantastic boost in an immunization protocol, activating the right B cells, but it may fail to induce the correct antibodies, which would cross-react to the native HIV gp120.

With this mind, using the same approach, as for eOD-GT6 multimerization, we designed a fully-glycosylated eOD 60-mer, carrying the native sequence (eOD-60mer). The GT6 design work showed that delition of the N276 glycan is critical to allow VRC01 germline binding; the same was recapitulated in the context of native gp120 (*105*). We tested the effect of N276 glycan deletion on interaction of eOD with several calculated germline antibodies. Upon deletion of N276 glycan, the eOD has extremely weak but detectable affinity for germline VRC01, NIH45-46 and PGV20 (Table 5.1). Hence, we produced two constructs eOD-60mer-N276D and eOD-60mer to see whether activation of germline is possible even with a non-germline-targeted eOD.

	Germline Ab Binding (μM)					
	GL-VRC01	GL-NIH45-46	GL-3BNC60	GL-CHA31	GL-12A12	GL-PGV20
eOD-N276D	**178**	**22**	NDB[a] (117)	NDB[a] (117)	NDB[a] (117)	**17**

[a] No detectable binding (highest concentration tested)

Table 5.1 – eOD-N276D binding to a panel of calculated germline anti-HIV bNAbs

5.1 Characterization of the eOD-60mer-N276D

The eOD-60mer-N276D was produced in 293S cells and purified by lectin affinity chromatography using the methods described in Chapter 4. The lectin-purified sample was further analyzed by SEC-MALS using a Superose 6 (GE Healthcare) column. The light scattering analysis confirmed that the substantial portion of the sample is composed of fully formed particles with an expected molecular weight of approximately 2,000,000 Da (Figure 5.1A).

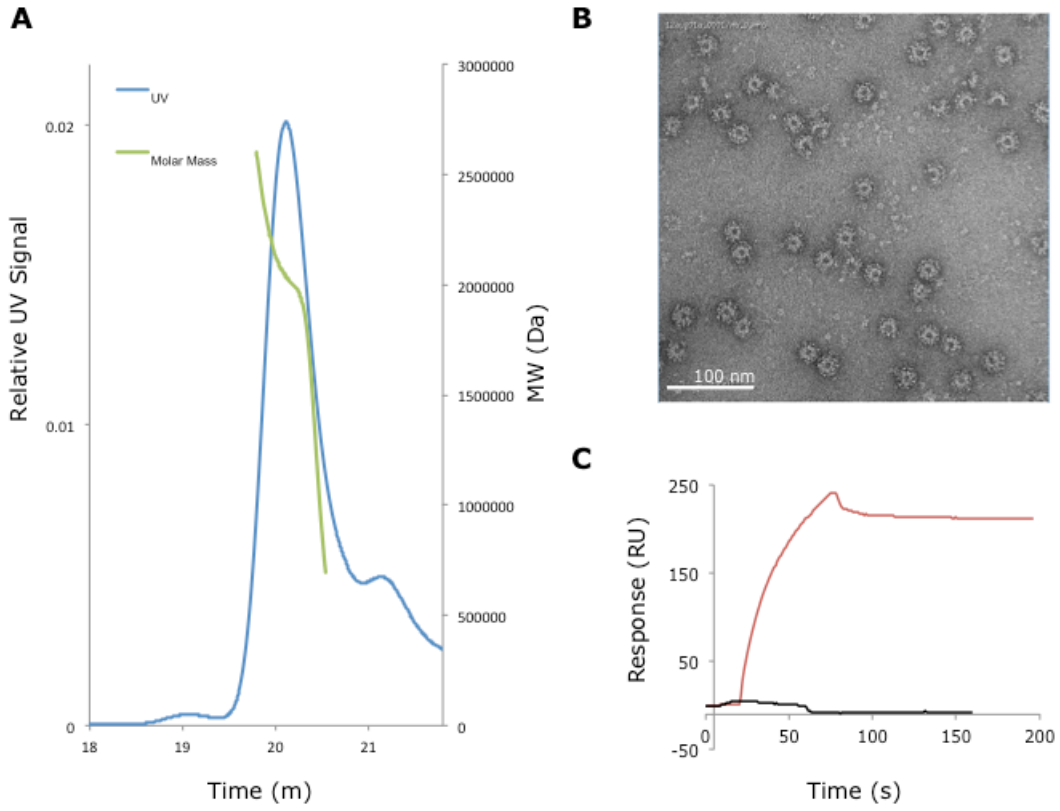


Figure 5.1 – Biophysical characterization of eOD-60mer-N276D: A) SEC-MALS analysis shows the major peak to be at the expected molecular weight of 2 x 10^6 Da (Blue – UV Signal, Green – Molecular weight); B) Negative stain image of the particles; fully and partially formed particles are present in the sample; C) SPR curve response for eOD-60mer-N276D (red, upper curve) and eOD-60mer for germline VRC01 ligand

The distribution of assembly states for this construct is different from the eOD-GT6 60mer. It appears that these 60-mers are not as efficient at forming complete 60 subunit complexes. We have observed this phenomena in other eOD variants and it appears to be construct and sequence specific, suggesting that it can be due to different processing or possibly different glycosylation efficiency of certain constructs. We are working to understand the factors affecting the efficiency of assembly. In the meantime, Superose 6 purification is capable of separating fully formed from partially assembled particles in order to get a homogeneous sample. Once purified, the particles remain stably assembled for weeks.

Consistent with the SEC-MALS analysis, negative stain images of the lectin purified samples show a number of fully formed and partially assembled species (Fig. 5.1B). The particles in the images are of the correct size, approximately 30 nm in diameter, and appear to be decorated with small molecules on the surface, presumably eOD.

To confirm the presence of eOD on the surface of the purified particles, we performed a binding experiment by SPR. Germline VRC01 was captured on the chip surface and eOD-60mer and eOD-60mer-N276D was flowed as analyte. As expected, eOD-60mer-N276D bound irreversibly, whereas, eOD-60mer did not respond at all (Fig. 5.1C). Therefore, the eOD-60mer-N276D particles follow the same binding trends for the calculated germline antibodies as the monomer (Table 5.1) and the full-length gp120 (*105*).

5.2 In vitro B Cell activation

Using two available in vitro B cell systems, one displaying NIH45-46 (*92*) IgG and another VRC01 (*91*) IgM, we carried out calcium flux experiments, as described in the previous chapter. B cells displaying IgGs are known to activate more readily than IgM-displaying B

cells and are considered to be a less stringent test of biologically relevant activation. At the time, we did not have complementary pairs of either NIH45-46 or VRC01 in both IgM and IgG format; hence, we present data from two different systems using two different antibodies, generated in two different labs. Both eOD-60mer-N276D and eOD-60mer readily activate mature versions of NIH45-46 and VRC01 at 100 nM and can be titrated down even lower (Fig. 5.2).

eOD-60mer does not bind germline VRC01 in SPR (Fig. 5.1C) and does not activate B cells expressing germline VRC01 IgM, even at 27 μM concentration (Fig. 5.2; black line). In the case of NIH45-46, activation is dramatically reduced, as compared to eOD-60mer-N276 at 10 times lower concentration (500 nM, blue line). eOD-60mer-N276D readily activates germline NIH45-46 at 1 μM and 500 μM. The activation signal for germline VRC01 IgG is very weak. The experiment was repeated in triplicate one week apart with different batches of cells. eOD-N276D-60mer specifically activates germline VRC01 IgG expressing B cells at a concentration of 100 nM.

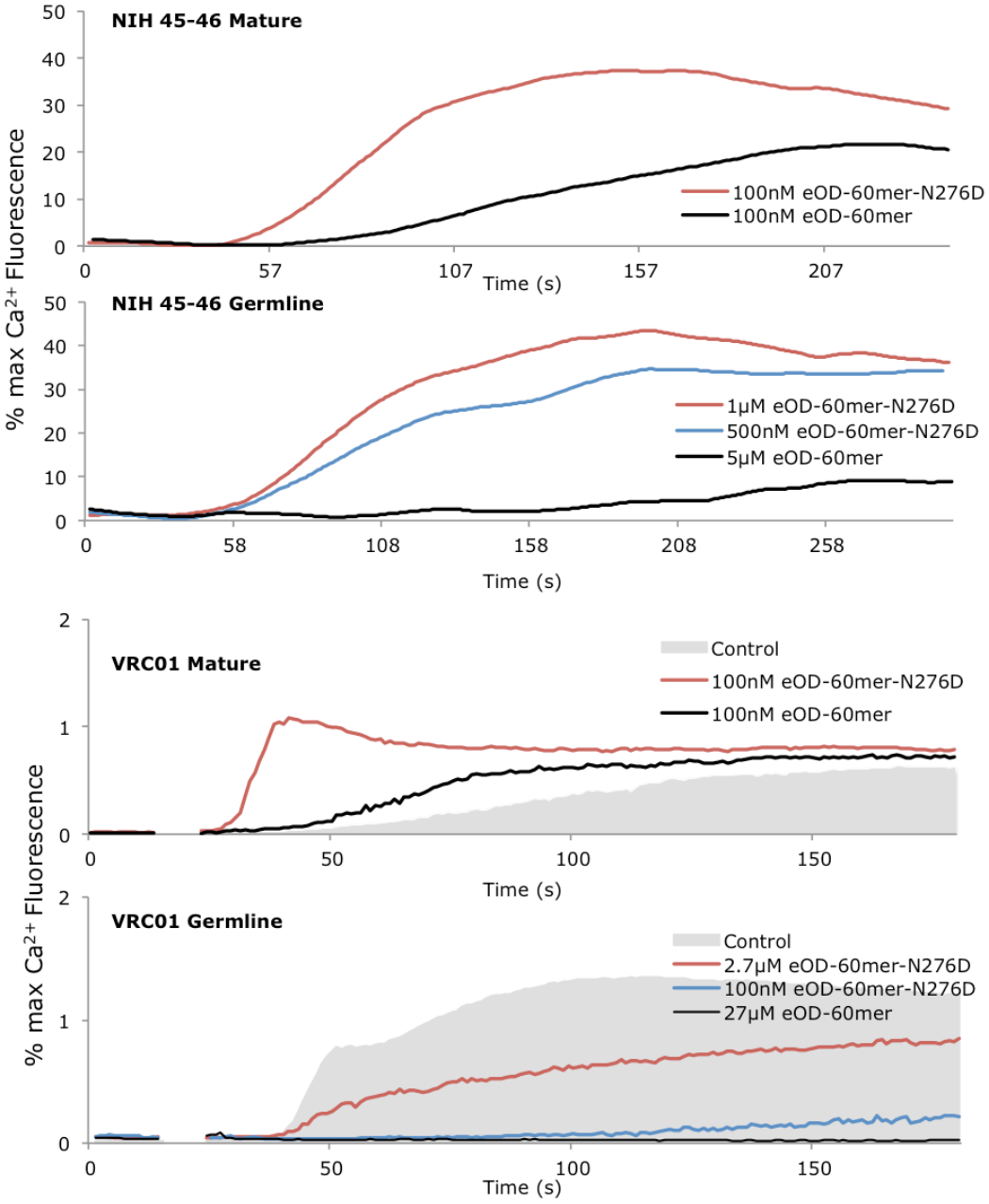


Figure 5.2 – In vitro B cell activation: Calcium flux experiments of eOD-60mer-N276D and eOD-60mer activating B cells displaying NIH45-46 IgG and VRC01 IgM, germline or mature.

5.3 Discussion

Contrary to the previous chapter, the eOD construct on the surface of the 60-subunit nanoparticle described in this chapter carries the native sequence of HIV gp120 and is fully-glycosylated, except for the N276D glycan, deletion of which is required for binding germline VRC01. The high degree of avidity afforded through multimerization is capable of surmounting the low affinity of eOD monomer to calculated germline VRC01 and other bNAbs. Hence, we now have a native-like eOD, which readily activates mature and germline versions of a key broadly neutralizing anti-HIV antibody. Thus, the eOD-60mer-N276D construct is an attractive immunogen, which is, in theory, capable of activating appropriate germline B cells while providing the native HIV gp120 surface for subsequent affinity maturation.

The examples presented in Chapter 4 and 5 demonstrate the potential benefits of multimerization. While the constructs used in the two chapters are only marginally different from each other, we are presently working to display other immunogens, including HCV1 epitope scaffolds, on the surface of the 60mer to prove the general applicability of this platform.

CHAPTER 6: B CELL EPITOPE PREDICTION

In Chapter 2, I presented the development of epitope scaffolds, which elicit neutralizing HCV antibodies in mice. In Chapter 4, I described a multivalent presentation platform, which we hope to use to increase the immunogenicity of our scaffolds in animal studies. Looking downstream, our next goal was to improve the methods available for the analysis of animal and human sera responses testing immunogens such as the epitope scaffolds described in Chapter 2 and minimal constructs from Chapter 4 and 5. Additionally, we were looking for opportunities to evaluate the epitopes, which our immunogens are presenting, in an effort to reduce off-target epitopes prior to immunization trials. In order to answer these questions, we developed a novel computational epitope identification protocol, competent to work with heavy glycosylation and flexibility of HCV and HIV proteins.

6.1 Challenges in B Cell Epitope Prediction: An HIV-1 Perspective

B cell epitope usage (BCEU), most commonly presented in literature as antibody epitope usage, may be defined as preferential targeting of a subset of available surfaces on a given antigen, leading to a skewed immune response towards a limited set of potential epitopes. Such differences in BCEU are most vividly exemplified by so-called immunodominant epitopes, which are targeted by a significant fraction of all antibodies produced by the immune system in response to a particular antigen. In the viral world, the sequence of immunodominant regions is often hypervariable, the variable loops of HIV glycoprotein gp120 being a perfect example, which is an evolutionary mechanism of directing immune response toward non-essential regions rendering key functional sites immunosilent. Computational BCEU (cBCEU) prediction offers a rapid, low cost and effort alternative to experimental antibody epitope identification.

In theory, reliable cBCEU prediction would allow identification of non-essential immunodominant regions removal of which should focus the immune response to more relevant epitopes. Additionally, the epitopes identified could be used as vaccine targets or targeting of therapeutic antibodies (*62*). However, the structural heterogeneity of B cell receptors (BCRs), which forms the basis for antibody recognition, the intrinsically random selection process by which B cell clones compete for proliferative signaling and general lack of understanding of the rules governing immunodominance has, to date, rendered cBCEU prediction per se intractable at a level useful for vaccine development. Current best-practice protocols in cBCEU prediction typically use solvent exposure of antigen residues (CEP (*63*)), often adding statistical information derived from non-structural experiments (DiscoTope (*64*), BePRO/PEPITO (*67*)); antibody shape complementarity using representative antibody probes offers additional information (Ellipro (*66*)). Even the best cBCEU prediction methods achieve positive identification accuracies below 30% (*106*), likely due to the plethora of theoretically possible epitopes in contrast to the limited set remaining after B cell selection.

The low success rate of cBCEU prediction methods is daunting, but there are numerous opportunities for improvement. From a structural perspective, all of the presently available structure-based cBCEU prediction methods view the molecule as a static, rigid object. This is clearly an oversimplification, as we know that a protein is flexible in solution and will sample a number of conformations over time, granted that some of them are preferred more than others. In the case of HIV gp120, a rough alignment of several crystal structures will reveal a relatively conformationally stable core surrounded by flexible regions. These regions will hide and expose epitopes over time, by sampling alternate conformations in solution. Ignoring such motions, as most cBCEU prediction methods do, diminishes the biological significance of the predictions and most likely affects their accuracy. Advances in protein prediction, specifically in the context of the Rosetta protein modeling package (*107*), allows

accurate modeling of protein structure for small segments of protein. Using Rosetta, it should be possible to simulate the conformational flexibility of surface exposed regions and improve the biological relevance of structure-based cBCEU prediction.

A second oversight, which is not addressed by any of the current cBCEU prediction methods, is viral protein glycosylation. Proteins on the surface of HIV, HCV and influenza are heavily decorated by self-sugars acquired during the viral replication life cycle. While glycosylation does eliminate immune response for a given region entirely, it will significantly dampen it (*108*, *109*). Hence, it is critical to take antigen glycosylation into consideration when generating biologically relevant BCEU predictions.

The goal of the work contained herein is to address the described limitations of cBCEU methods and illustrate the advantages by applying the new method to HIV-1 gp120 protein, which is a perfect candidate due to its high conformational flexibility and extensive glycosylation.

6.2 N-Linked Glycosylation Masks Large Areas of HIV-1 gp120

A common feature of the eukaryotic secretome is post- or cotranslational modification of proteins by the covalent addition oligosaccharide groups. N-linked glycosylation chemically modifies surface asparagine residues found in an N-X-(S/T) motif, where X is any non-proline amino acid and (S/T) either a serine or threonine residue. Subsequent enzymatic modification of the N-linked glycan allows tailoring of, for example, antibody effector functions (*110*); the process of enzymatic glycan modification results in a heterogeneous population of glycoprotein species sharing an identical protein sequence.

While antibodies against non-human glycan structures have been observed to play an important protective role against parasitic eukaryotic pathogens, viral proteins display human glycans as a consequence of hijacking host intracellular machinery. As such, B cells targeting most viral glycan-containing epitopes fail to escape immune tolerance mechanisms, partially shielding viral surface glycoproteins from adaptive immune responses.

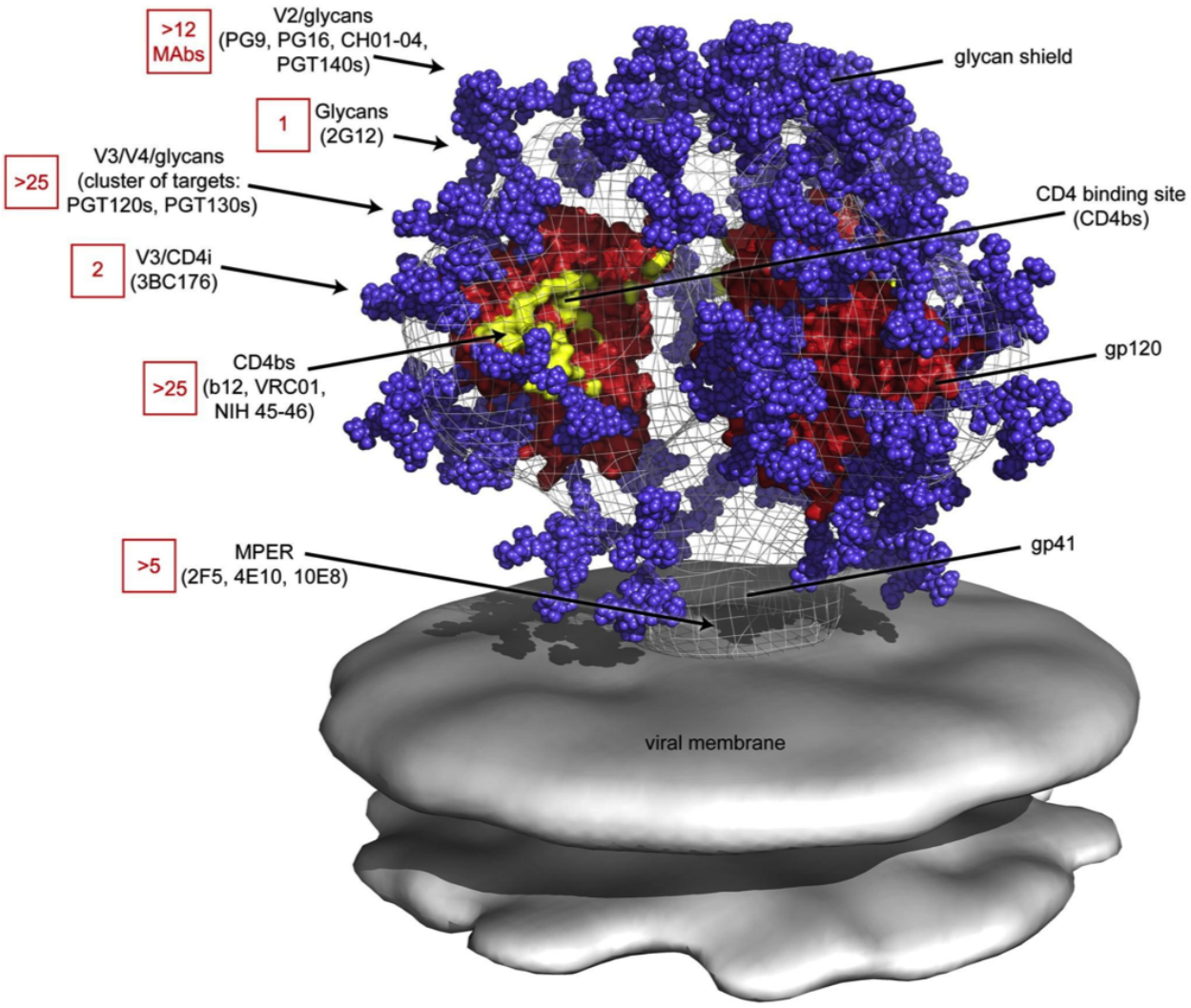


Figure 6.1: Glycosylated gp120 Models Docked into a Cryo-Electron Microscopy Structure of the HIV *Env* Trimer. Epitopes used by broadly neutralizing antibodies are noted; glycans corresponding to missing V1/V2 loop density are placed above trimer cap. Figure is reprinted with permission from Elsevier Press (*17*).

The amino acid sequence of HIV-1 gp120 contains more than 20 hypothetical N-linked glycosylation sites (*111*). gp120 together with gp41 forms a non-covalently linked homotrimer on the surface of the HIV virion. In the virion the bulk of the glycans reside on the outward face of the gp120 subunit, known as the outer domain (OD) (Fig. 6.1). Consistent with this observation, immunization with gp120 monomer will produce few antibodies to the densely glycosylated OD. Instead, the bulk of the response will be against the inner domain, the inward facing region of gp120 of the surface trimer, which is devoid of glycans, and immunodominant variable loops.

However, this is not to say that antibodies are not produced against the glycosylated OD. In rare cases, the immune system is able to penetrate the glycan shield and identify relatively unglycosylated patches on the protein surface. Several such antibodies have been identified; some are broadly neutralizing and recognize epitopes on the OD and base of V3 and V1/V2 loops (Fig. 6.1).

Antibodies against the human glycans on gp120 themselves are rare. 2G12 is the only known broadly neutralizing antibody (bNAb) directed against an uncommonly dense high mannose patch.

In this work, we seek to capture the effect of glycan masking and discard theoretical epitopes, which would otherwise be considered if glycosylation is to be ignored. By simulating glycan and underlying backbone flexibility, we aim to identify deglycosylated protein patches nestled between glycans, similar to the bNAb epitopes on the OD. Finally, this work brings the field closer to modeling and identifying glycan only epitopes, granted that further development of methods accurately describing glycan-protein interactions needs to be carried out.

6.3 Vaccine Sieve Analyses

Determining the degree of accessibility of potentially occluded epitopes is of great interest to vaccine design, particularly in the case of heavily glycosylated viral envelope proteins such as the HIV-1 Env protein and influenza hemagglutinin trimers. While we expect the method discussed herein to have considerable utility in evaluating potential vaccine components, the initial motivation for the development of this method was the need to identify antibody-accessible antigen positions to aid the sieve analysis of the RV144 HIV vaccine trial (*69*).

The RV144 trial represents an important milestone for the HIV vaccine field in that, for the first time, modest protection was observed in the vaccine cohort; the protection was significant ($p < 0.05$) when counting patients who completed the vaccine regimen, but not when counting by intent-to-treat, due to less observed protection in the incompletely vaccinated subgroup. Extensive follow-up research has attempted to disentangle potential mechanisms of protection in RV144; a particularly useful tool for post-hoc analysis of the vaccination response is the identification of statistical correlates of protection among parametrizable aspects of the trial subjects' immune responses. IgG response against the gp120 V1/V2 loop emerged as a significant correlate of protection by analysis of the vaccinated and unvaccinated infected trial subjects; IgA response against gp120 was a negative correlate of protection.

Immune correlates of protection are not, in themselves, sufficient to demonstrate a protective response to vaccination; a point of particular importance is the possibility that a given immune response parameter may offer protection against viral infection due to a vaccine-independent variability in the trial subject population. Sieve analysis examines the viral sequences obtained from infected trial subjects to determine selection effects of

vaccination on potentially infective viral sequences; although mechanisms for such selection effects remain unclear, they may be causally attributed to the vaccination protocol.

Sieve effects, like other population-level genetic variability measurements, are intrinsically noisy; additionally, the signal-to-noise threshold required to attribute significance to a putative sieve effect at any individual sequence position decreases with the number of positions considered. As such, pre-filtering the set of analyzed positions without the use of information regarding infectivity or viral sequences increases the significance of observations regarding the smaller, post-filter set of target positions.

The work described in this chapter offers a method for performing such filtering, using computational glycoprotein structure modeling and optimized geometric models to rank protein sequence positions by accessibility to antibody binding.

6.4 Results

6.4.1 Rosetta Structural Prediction Generates Diverse Conformational Ensembles

To capture the inherent protein flexibility, we used Rosetta protein design package to model an ensemble of gp120 conformational states in order to understand the effect of intra-protein motions on accessibility of epitopes.

An alignment of existing crystal structures of gp120 reveals a well-conserved relatively in-flexible core decorated by flexible variable loops (V1/V2, V3, V4, V5) and N- and C-termini, which are often trimmed or invisible due to lack of density in the X-ray map. With a combination of RosettaRelax (*112*) and RosettaRemodel (*78*), we generated an ensemble of gp120 models for each RV144 vaccine component, which would theoretically span a wide

range of low-energy, favorable conformations of the antigen and would help us understand their effect on epitope accessibility. Over the whole ensemble, residues were as far as 50 Å from each other with the highest differences in the unconstrained N- and C- termini (Fig. 6.2).

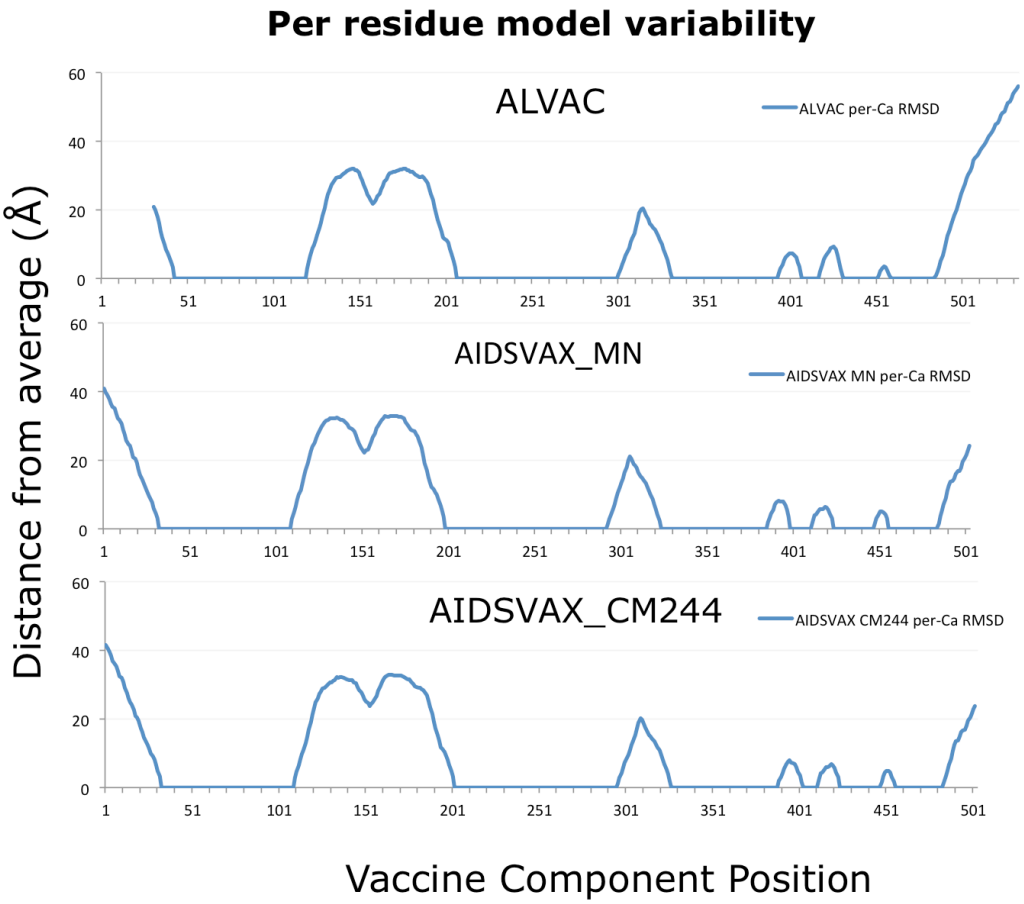


Figure 6.2: Model Variability vs. Amino Acid Sequence Position. Rosetta-built loops show variability in regions of sequence (blue) corresponding to variable gp120 loops identified in prior studies. Regions accessible to antibodies frequently map to these positions, as well as the N-terminal region; flexible regions corresponding to the ALVAC C-terminal transmembrane helix and AIDSVAX N-terminal gD tag are particularly variable with respect to the conserved structural core, used for interstructure alignment here.

Each of the models from the ensemble was decorated with glycans using high-mannose core structures at each predicted N-linked glycosylation site, followed by Monte Carlo minimization of intersugar bond rotations to resolve any clashes.

The ALVAC component of the vaccine used in the RV144 trial carries a C-terminal transmembrane helix tethering the protein to the cell surface. This case is unique to RV144 and is not of general concern to soluble untethered antigens. The proximity of the membrane imposes additional filter on the biologically relevant orientations. To accurately simulate the range of motions of the membrane tethered ALVAC molecule, a second round of protein backbone modeling was performed, where a fully-glycosylated model, placed near the z=0 plane and linked to a transmembrane helix, was allowed to explore alternative placements in space as the linker, spanning the region between the membrane and the core of the molecule, was allowed to sample alternative conformations in RosettaRemodel. Any linker conformation, which placed the ALVAC molecule below the z=0 plane was discarded.

6.4.2 Atom-Centered Spatial Fab Orientation for Epitope Patch Prediction

As antibody complementarity-determining regions (CDRs) vary considerably in both length and sequence, and as full-atom energetic evaluation of protein-protein interfaces is both computationally expensive and imprecise, the EPIMAP method focuses on the determination of accessibility for sites on a target glycoprotein's surface using a sphere as a proxy to represent the complexity of the CDRs. This rests on the assumption that at sufficiently close antibody placement some combination of CDRs will be capable of forming an energetically favorable, high-affinity interface with any "patch" of glycoprotein surface residues, assuming that any steric clashes can be resolved by flexibility of the glycans and predefined protein regions (Fig. 6.3, Fig. 6.4).

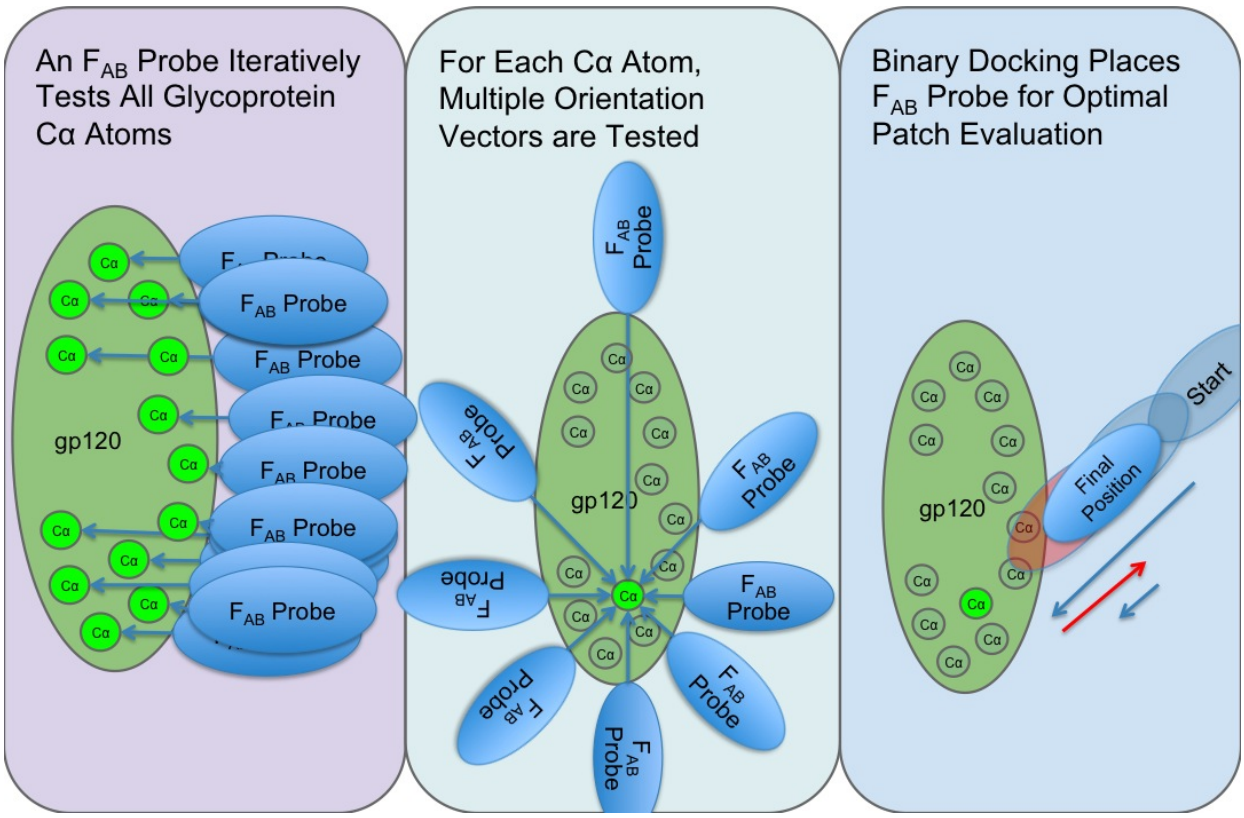


Figure 6.3: Schematic Representation of Antibody Placement Procedures. A F_{AB} probe model is placed in multiple orientations (center panel) surrounding each Cα of the target glycoprotein model (left panel); optimal positioning of each probe orientation is achieved by a binary-search docking procedure (right panel) with the probe making successively smaller movements towards the glycoprotein without detected clashes (blue) or away from it when clashes are detected (red).

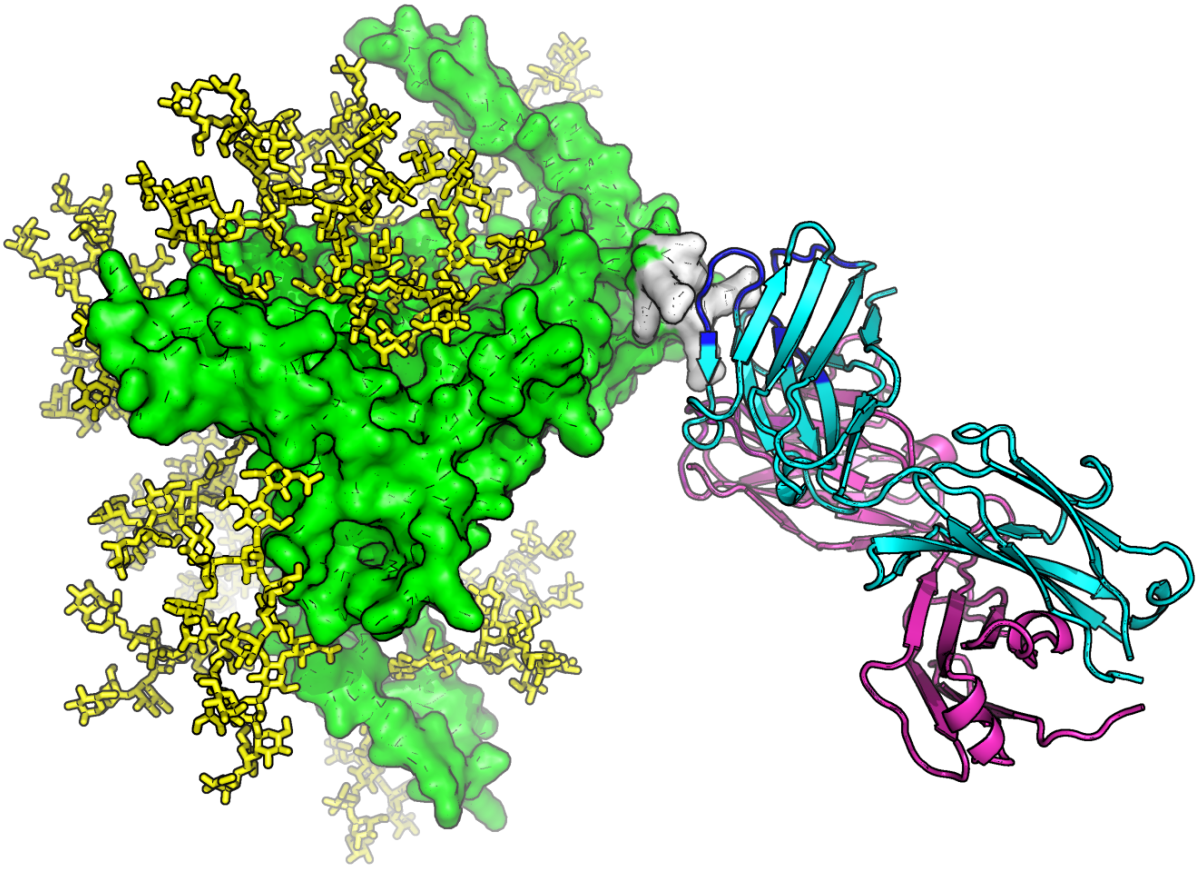


Figure 6.4: Antibody Patch Definition. The 3HMX antibody probe (cyan heavy chain, magenta light chain) is positioned adjacently to the gp120 (green) surface; a "patch," (white) defined as all glycoprotein atoms within 6Å of a designated interaction subset of the antibody corresponding to the antibody CDRs (blue heavy chain, purple light chain). No glycans (yellow) participate in this patch.

To generate potential antibody placements, a full-atom "query" model of an antibody binding fragment (Fab) is iteratively placed in close proximity to each Cα atom of a glycoprotein model of interest (Fig. 6.3, left-hand panel); this process is repeated over a set of randomly oriented target models incorporating modeled structural variability for both protein and glycan components. Antibody placement covers all spatial approaches to each target model atom by alignment of the query model to an approach vector (Fig. 6.3, center panel), and subsequent rotation about that vector. The depth of sampling for a particular target model can thus be increased by the use of additional approach orientations and rotational operations performed about the approach vectors. Proximity to the target glycoprotein is allowed by the use of a binary search procedure along the approach vector to determine a closest distance without steric clashes (Fig. 6.3, right-hand panel)

Evaluation of antibody placements consists of an initial steric clash check to rule out physically impossible interaction models, and subsequent identification of hypothetical epitope residues on the target glycoprotein selected by proximity to the CDR region of the Fab probe. A key performance improvement of this version of the code is the use of a spatially indexed binning system for the viral glycoprotein model; this allows steric checks, which are implemented as minimum distance cutoffs, to be applied only to nearby atoms in the Fab probe. Since the Fab model contains approximately an order of magnitude fewer atoms than the glycoprotein model, spatial indexing of the glycoprotein enables computationally expensive geometric operations to be performed only on the smaller query model without reinitializing or updating glycoprotein atom positions.

6.4.3 Fab Models Are Interchangeable for Patch Accessibility Prediction

Initial patch prediction for the RV144 sieve analysis was performed on a glycosylated gp120 model using a structural model of the HIV bNAb b12 (*36*) Fab construct; this provided

recapitulation of the known b12 epitope as a method viability metric intrinsic to the data set. Subsequently, we questioned whether b12 was the best choice for a general probe for patch prediction. We developed a benchmark set of human high-resolution Fab:antigen complex crystal structures to test alternative Fab probes (Appendix - Table 6.1).

Using the benchmark set of human Fab:antigen complexes, we asked the question – "Using our protocol what is the maximum proportion of the epitope that we can identify within a single patch?". Note, our protocol does not unique identify the epitope; it will only report a patch, which will contain the epitope. In theory, the native Fab probe has the best possible shape complementarity for the epitope and the value reported is our theoretical best case. However, this is not true, as the method is coarse-grained and the limited set of approach orientations examined may be incapable of recapitulating the native mode of binding. Hence, it is possible for a general Fab probe to outperform the wild-type probe in this experiment. Comparison of epitope recapitulation with either b12 or the native Fab probe for each target complex revealed comparable performance between our surrogate and native models, lending further support our initial results (*69*) obtained using the b12 probe. For the cases where b12's performance differed from the native probe performance significantly (15% or greater loss of recapitulation) there was a clear trend towards shorter CDRH3s (Fig. 6.5A). A possible explanation for this is inability of b12 to place sufficient CDR residues within the proximity cutoff of a target epitope required for proper scoring due to a steric clash by the probe's longer CDRH3. In contrast, Fab models with shorter CDRH3s will tolerate greater proximity to the antigen and will be placed in a correct orientation more often. The downside of short CDRH3 probes is their inability to identify residues buried in loop-accessible cavities within a larger epitope, for example, in cases such as 3UY7 and 3HI1.

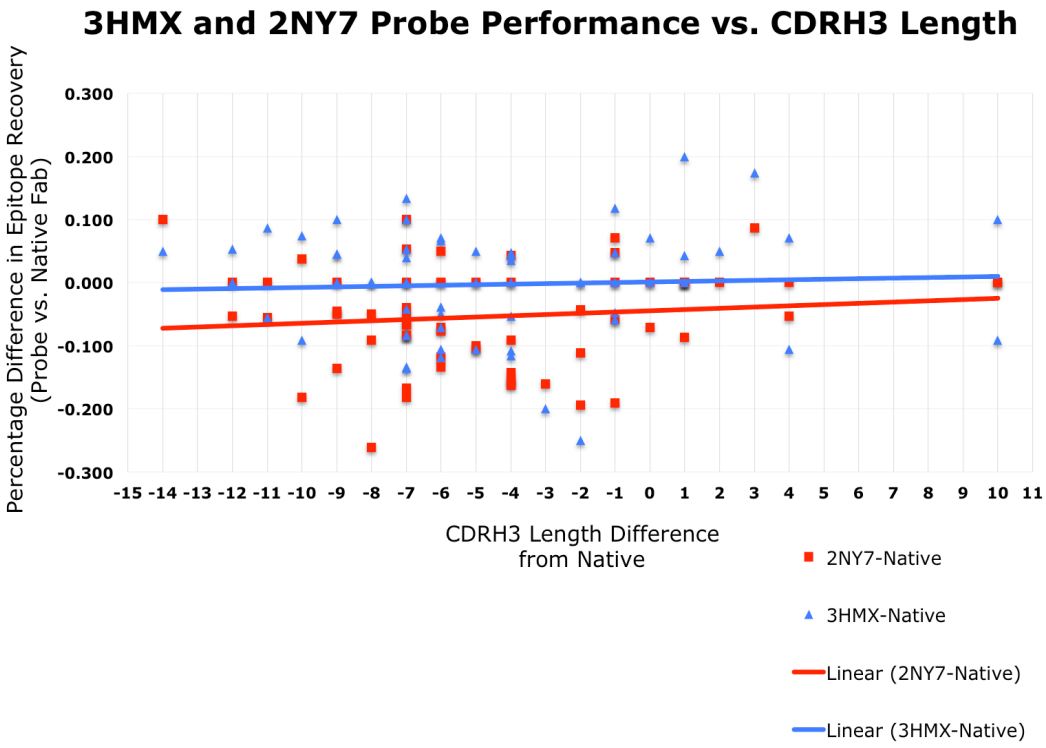


Figure 6.5A: Bechmark performance for 3HMX and 2NY7 Probes – Summary of difference of epitope recovery using the native or 3HMX (blue)/2NY7(red) probe for each complex plotted versus length difference of the CDRH3 loop of the probe compared to the native antibody for the given complex. Linear trendline was generated by fitting a line to all datapoints for 3HMX (blue) and 2NY7 (red). 3HMX probe outperforms 2NY7 probe by a small margin.

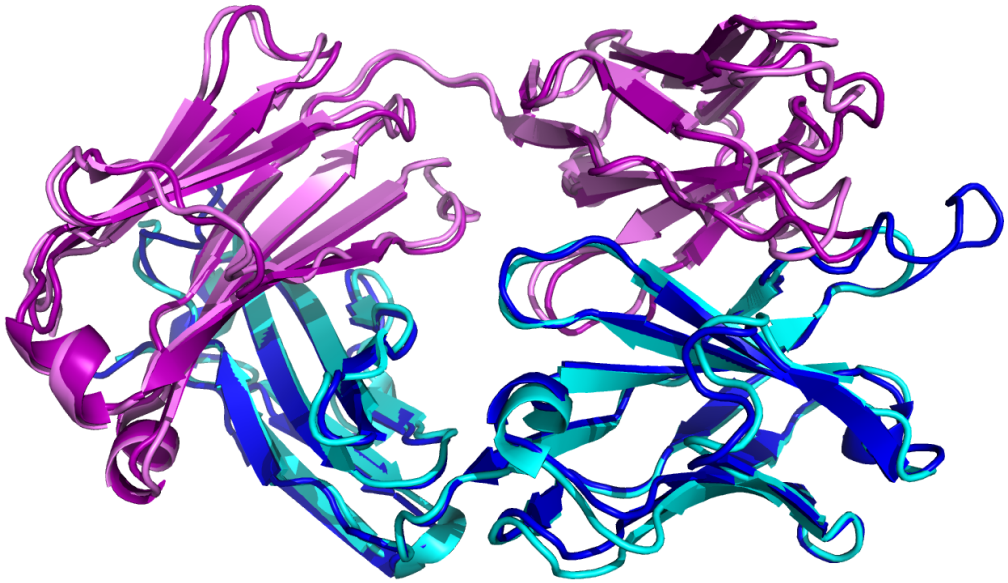


Figure 6.5B: Structural Alignment of the 3HMX and 2NY7 Fab Models Used as Probes. 2NY7 (darker blue and purple for heavy and light chains, respectively) has a comparatively long CDRH3 loop, which protrudes past that of the 3HMX structure (cyan and violet for heavy and light chains) on the right-hand side of the figure.

In order to make our method generally applicable, with the help of the benchmark, we chose to use Fab probe based on the ustekinumab antibody (PDB ID: 3HMX), which performed exceptionally over the whole benchmark and appeared to be a middle-ground compromise (Appendix – Table 6.1). We note that it may be prudent to choose an alternative Fab probe for applications to mapping antibody accessibility in other organisms. It is also clear that increasing the number of orientations, at the cost of computational run time, can improve the recovery of the full epitope. In this work, we chose to use the 3HMX Fab probe and limited set of orientations in order to maximize both performance and accuracy of epitope identification.

6.4.4 Glycosylation Sterically Hinders the Majority of Surface Epitopes

A key motivation for the development of the EPIMAP method is our assumption that glycosylation of antigen surfaces will broadly attenuate full-atom Fab binding. To test this

concept, we generated patches for fully glycosylated and deglycosylated versions of the same set of 1000 gp120 models, for each component, and compared patch counts across the two sets.

Comparison of the accessible surface patches between glycosylated and deglycosylated models reveals that, as expected, patches detected by the EPIMAP algorithm are dramatically reduced on a per-residue basis in the presence of glycosylation (Fig. 6.6). Overall, in the absence of glycosylation, EPIMAP identifies patches permissively across all surface positions with minimal discrimination between positions. However, when glycosylation is considered, the number of potential residues in patches is dramatically reduced and a clear set of well-defined regions emerges.

Notably, while certain areas of the glycoprotein were not affected by any significant steric hindrance from nearby glycans, the majority of surface-exposed residues found on the deglycosylated models were. Finally, several regions falsely predicted to be accessible, in the deglycosylated models, are in fact entirely obscured, in the glycosylated models.

Regions of high exposure such as the V2 and V3 loops of gp120, which can be accessed easily bypassing the glycans especially in the monomer, are prominent in both glycosylated and deglycosylated model predictions. Removal of glycans reveals additional angles of approach, as supported by elevated patch counts determined from the deglycosylated set.

In light of this result, it is clear that structure-based cBCEU prediction methods, which do not account for glycosylation will have a difficult time discriminating between true epitopes present on the antigen, as illustrated by EPIMAP results from the deglycosylated protein set, which predict epitopes uniformly across the entire sequence even if the region is decorated by glycans and is most-like inaccessible. Modeling glycans as presented in our method

eliminates a large number false-positive epitopes and can improve the performance of any downstream prediction method.

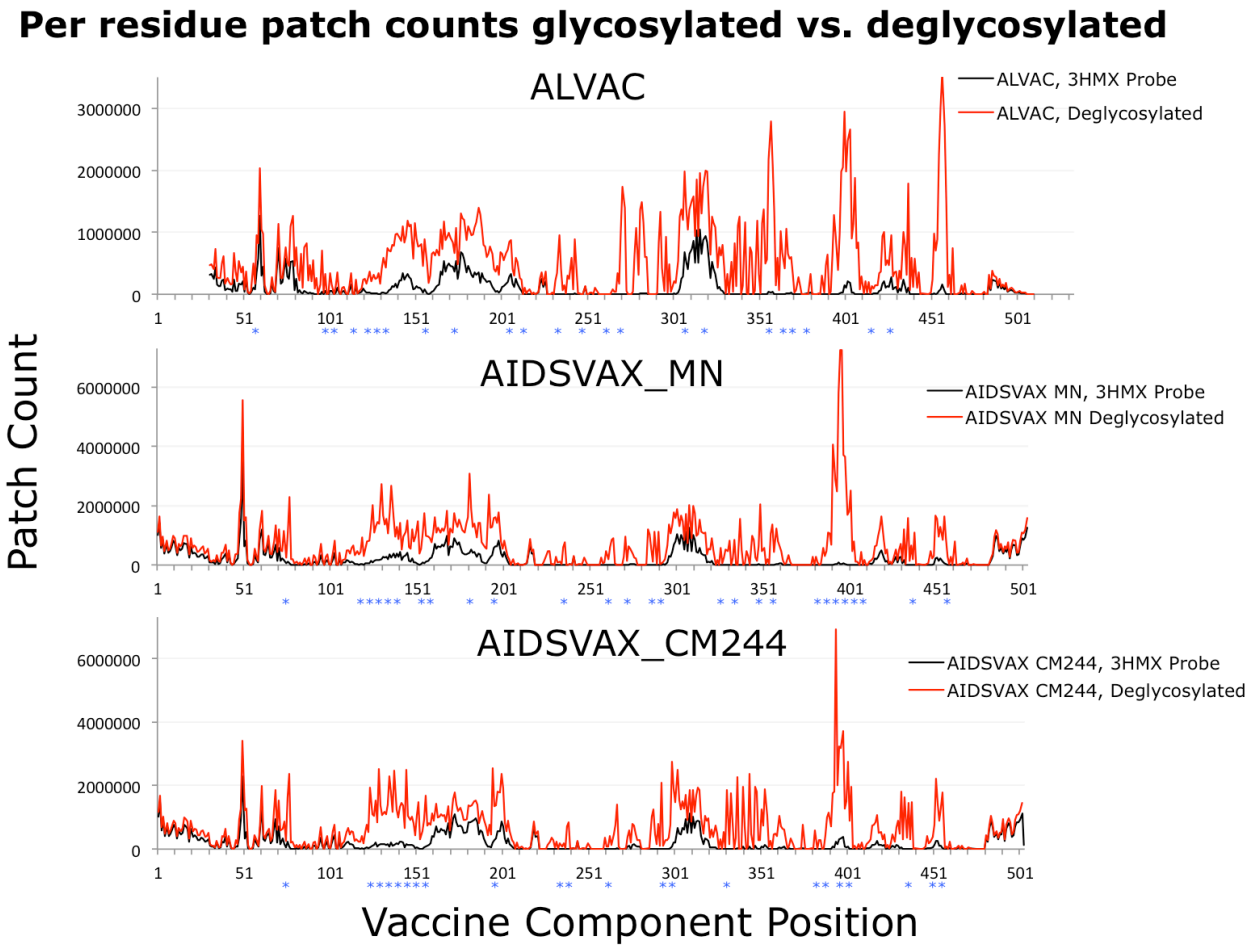


Figure 6.6: Glycosylation-Induced Reductions in Epitope Accessibility. Deglycosylated models' patch count outputs (red lines) are uniformly higher than those of the fully glycosylated models (black lines); regions particularly well-shielded by glycans include the V1 loop (residues 120-151) and the V4 loop (residues 390-400). The putative glycosylation motifs for each sequence are denoted by a blue asterisk.

6.4.5 EPIMAP Analysis of Computationally Modeled RV144 Components Matches Microarray Immunogenicity

EPIMAP provides a ranking of hypothetical epitope accessibility on flexible, glycosylated proteins. Ideally, experimental validation tools would provide similar epitope statistics to make the comparison fair. However, BCEU is a complicated, poorly understood process, which is not based entirely on epitope accessibility. Nevertheless, it is important to verify that our predictions translate, to some degree, to actual epitope usage observed in sera against the antigen in question.

In the context of the RV144 post-analysis, sera from vaccinated individuals, exposed to all three vaccine components, was analyzed by assaying reactivity to a peptide microarray of overlapping fragments spanning the entire sequence of each component. The peptide reactivity profiles reveal which linear epitopes in the sequence were immunogenic. Note that microarray will only detect linear epitopes, whereas, EPIMAP should be able to capture both linear and conformational epitopes.

From a high-level perspective, EPIMAP predictions of epitope accessibility track well with the microarray sera response (Fig. 6.7). There are several additional regions, which are predicted to be accessible by EPIMAP, but are not immunogenic according to the microarray data. These can be grouped into three parts 1) V5 (residues 451 – 461), VD (residues 275 – 285) and Loop E (residues 351 – 361) 2) Bridging sheet (residues 201 – 211, 421 – 441) 3) N- termini and C-termini, in the case of AIDSVAX_MN and AIDSVAX_CM244. Group 1 and 2 represent conformational epitopes, which are frequently sampled by EPIMAP but cannot be detected by linear peptide microarray. Group 3, N- and C-termini, are extended, relatively unconstrained regions of the protein, which probably lack secondary structure. EPIMAP will

frequently sample these positions, but the reasons for their poor immunogenicity in-vivo are unclear.

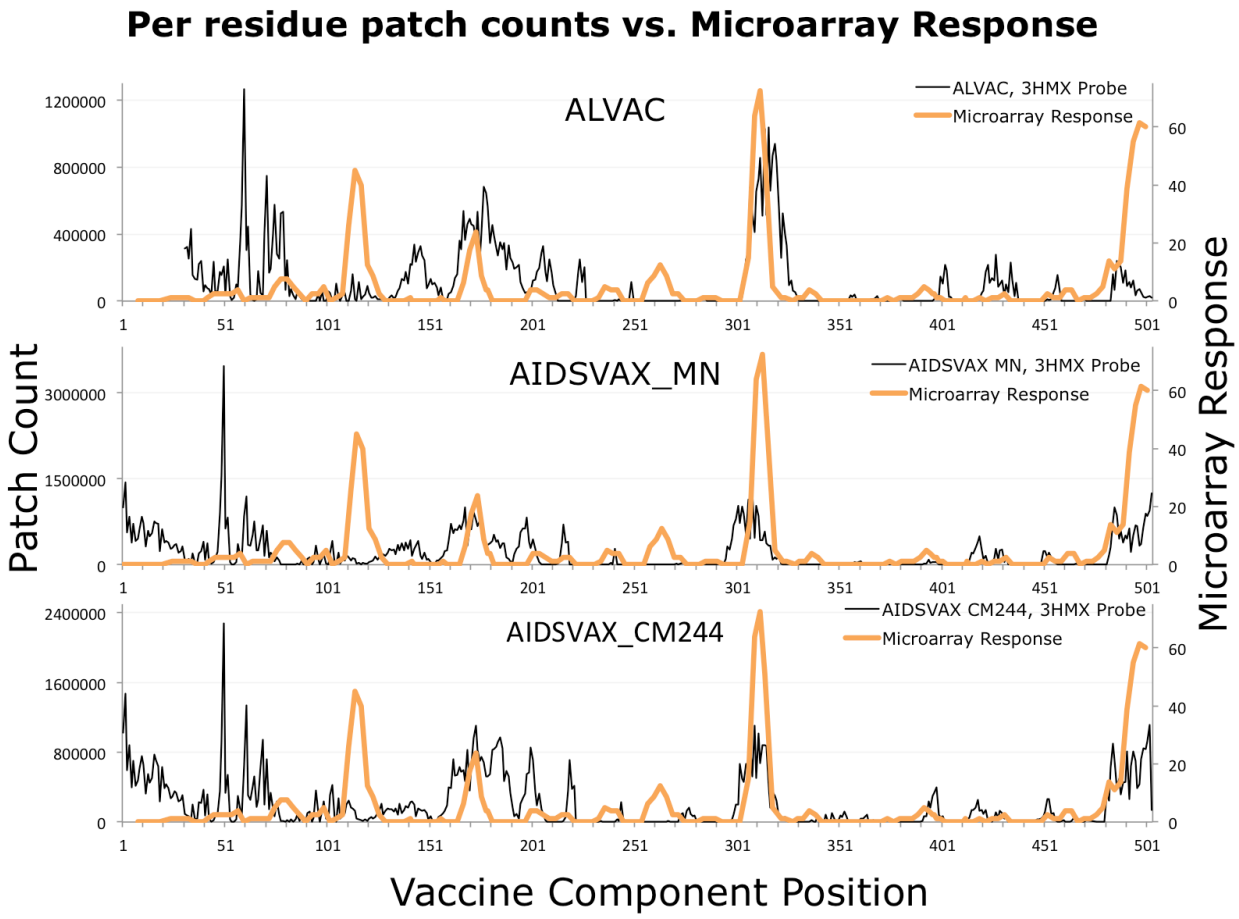


Figure 6.7: Microarray Experimental Evaluation of EPIMAP modeling. Microarray detection of 15mer peptides spanning the sequence of gp120 (orange) largely tracks with EPIMAP predictions (black); exceptions are in regions, which are inaccessible in our model and were not modeled as flexible, i.e. 101-117 (Fig. 6.2) and conformational epitopes, which are not detectable by microarray.

There are several major discrepancies between EPIMAP-predicted accessible regions and observed antibody binding sites observed in microarray. The first occurs in gp120 residues 120-130; this region, comprising a short stretch of the bridging sheet at the base of the V1/V2 loop (*113*), is known to be highly flexible in its native viral context. However, it was not obvious from the crystal structures that we should explore flexibility in the region. The second region, loop A, (residues 231-241) nestled between glycans at positions 87, 237, 244 and 279 was also not modeled in our conformational ensemble as it appeared to be a well-packed, stable region. Modeling complete protein motions would be an ideal solution to resolve this problem. At present, we can only reliably model short stretches of protein, so there will always be regions, which can potentially be substantially fluid but are overlooked by our modeling methods. We recognize that future work should focus on resolving the limitations of our conformational ensemble building procedures. This is especially obvious in the case residues 255-275, which are strongly immunogenic in microarray and silent by EPIMAP predictions. This stretch is contained entirely in the core and is not solvent accessible. In the absence of protein-wide motion analysis it is impossible to identify this region as a potential epitope.

Of the three viral sequences assayed, EPIMAP accessibility predictions for the CM244 sequence agree with microarray assay results the most. CM244 antigen elicited the most antibody response of any of the three vaccine components (*114*); thus, we theorize that decreased similarity between EPIMAP predictions for ALVAC and AIDSVAX_MN and microarray results is partially due to the fact that the bulk of the antibodies are less like the antibodies expected in response to either ALVAC or AIDSVAX_MN antigen. It would be possible to test this theory if sera was available for subjects exposed only to one antigen instead of a mix of three components.

Comparison of antigenic regions detected by the microarray experiment and predicted accessible regions from EPIMAP reveals strong correlation, but not ranked-order matching, between experimentally and computationally identified residues of interest on the surface of gp120 when considered as a whole. As EPIMAP does not consider epitope antigenicity per se, this result represents a successful prediction in that antigenicity is not detected in putative blocked regions; the experimentally demonstrated antigenicity of low-exposure regions may also usefully assist us in setting a threshold of accessibility for EPIMAP results to identify residues, which are sufficiently exposed to effectively recruit B cell responses.

6.5 Discussion

Given the pressing need for analyzing the immune response against heavily glycosylated viral proteins such as HIV gp120, the EPIMAP method presented here offers a useful insight for pre-filtering potentially antigenic epitopes for consideration within the context of a larger study. As previously stated in the introduction, current best practice in computational B cell epitope prediction is insufficient in itself to rank potential epitopes' antigenicity, particularly for large, flexible targets such as gp120.

Our results for the differential analysis of glycosylated vs. deglycosylated models vividly illustrate the importance of considering glycan-masking effects in any cBCEU prediction. Ignoring glycosylation muddles the analysis with false positive epitopes, thereby, reducing their accuracy.

Predictions from EPIMAP are in agreement with collaborators' microarray experiments for linear epitopes. The three regions, which we failed to identify, are either buried or occluded by nearby glycans. EPIMAP fails to identify all experimentally determined antigenic regions of gp120 as a direct result of sampling of a limited set of protein regions in the model

building step of our protocol, particularly in light of the data regarding glycan-mediated exposure and the boundaries of backbone flexibility selected for initial model building. For example, the 110-120 sequence region lies at the outer edge of the extended V1/V2 flexible zone, expanding the boundaries of the modeled region would have rendered our structural sampling problems intractable on an acceptable time scale using available hardware. Moving forward, we expect that improved structural prediction using experimentally derived constraints from the highly active field of HIV *Env*-focused structural biology should refine the results supplied by EPIMAP and related algorithms.

We envision further use for EPIMAP as a complementary method to experimental or computational B cell epitope identification methods, enabling the identification and filtering of otherwise antigenic epitopes, which are obscured by surrounding backbone and glycosylation. Further refinement of the computational methods should also focus on the parametrization of the patch interaction model; while our results largely recapitulate experimental observations, we have not exhaustively optimized our available parameter space. Ongoing publication of immunogenicity data for a variety of proposed vaccines should assist in this process by providing additional benchmark data sets.

The prohibitive time and financial costs of structural analysis for proposed glycoprotein vaccine components opens the door for methods such as EPIMAP to extend observed phenomena such as glycan-mediated epitope masking to untested protein sequences, providing a guide around such pitfalls in the design phase of candidate vaccines. Recent proof of principle for the rational design of glycan-mediated immune focusing into normally glycosylation-free proteins (*109*) offers another potential use for EPIMAP and related methods in vaccine design. We hope that open publication of the EPIMAP patch prediction code and test inputs will save other researchers time, effort, and money in moving their

own projects forward, and eagerly await refinements and improvements upon the methods presented here.

6.6 Materials and Methods

6.6.1 RV144 Vaccine Components

The RV144 vaccine trial used a recombinant canarypox vaccine (ALVAC-HIV) containing one gp120 sequence in DNA form followed by booster injections of recombinant gp120 protein (AIDSVAX) derived from two gp120 sequences (*115*), referred to here as CM244 and MN. The gp120 sequence encoded in the ALVAC component is modified by the addition of an C-terminal transmembrane helix (TM) intended to anchor the protein for cell surface display; the AIDSVAX components bear an N-terminal tag derived from the herpes simplex virus gD protein. Sequences for each vaccine component were kindly provided by Craig Margaret (FHCRC). For MN and CM244, we followed the *Berman et al.* definition (*116*) to assemble the gD-tagged sequence. Our models in the conformation ensemble carry the following sequences:

ALVAC with TM :
DNLWVTVYYGVPVWRDADTTLFCASDAKAQETEAHNVWATHACVPTDPNPQELHLENVTENFNMWKN
NMVEQMQEDVISLWDQSLKPCVKLTPLCVTLNCTNANVTNVKNITNVPNIIGNITDEVRNCSFNMTTEL
RDKKQKVHALFYKLDIVPIEDNTSSSEYRLINCNTSVIKQACPKISFDPIPIHYCTPAGYAILKCNDKNFN
GTGPCKNVSSVQCTHGIKPVVSTQLLLNGSLAEEEIIIRSENLTNNAKTIIVHLNKSVEINCTRPSNNTRT
SINIGPGQVFYRTGDIIGDIRKAYCEINGTKWNEVLKKVTKKLKEHFNNKTIIFQPPSGGDLEITMHHFNC
RGEFFYCNTTRLFNNTCMENETMEGCNGTIILPCKIKQIINMWQGAGQAMYAPPISGRINCVSNITGILL
TRDGGLNNTNETFRPGGGNIKDNWRSELYKYKVVQIEPLGIAPTRAKRRVVEREKRLFIMIVGGLVGLRI
VFAVLSVVNRVRQG

AIDSVAX_MN with gD tag:
KYALADASLKMADPNRFRGKDLPVLDQLLEVPVWKEATTTLFCASDAKAYDTEVHNVWATQACVPTDP
NPQEVELVNVTENFNMWKNNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLNCTDLRNTTNTNNSTANN
NSNSEGTIKGGEMKNCSFNITTSIRDKMQKEYALLYKLDIVSIDNDSTSYRLISCNTSVITQACPKISFEP
IPIHYCAPAGFAILKCNDKKFSGKGSCKNVSTVQCTHGIRPVVSTQLLLNGSLAEEEVVIRSENFTDNAK
TIIVHLNESVQINCTRPNYNKRKRIHIGPGRAFYTTKNIIGTIRQAHCNISRAKWNDTLRQIVSKLKEQFK
NKTIVFNQSSGGDPEIVMHSFNCGGEFFYCNTSPLFNSTWNGNNTWNNTTGSNNNITLQCKIKQIINM
WQEVGKAMYAPPIEGQIRCSSNITGLLLTRDGGKDTDTNDTEIFRPGGGDMRDNWRSELYKYKVVTIEP
LGVAPTKAKRRVVQREKR

AIDSVAX_CM244 with gD tag:
KYALADASLKMADPNRFRGKDLPVLDQLLEVPVWRDADTTLFCASDAKAHETEVHNVWATHACVPTDP
NPQEIDLENVTENFNMWKNNMVEQMQEDVISLWDQSLKPCVKLTPLCVTLHCTNANLTKANLTNVNNR
TNVSNIIGNITDEVRNCSFNMTTELRDKKQKVHALFYKLDIVPIEDNNDNSKYRLINCNTSVIKQACPKIS
FDPIPIHYCTPAGYAILKCNDKNFNGTGPCKNVSSVQCTHGIKPVVSTQLLLNGSLAEEEIIIRSEDLTNN
AKTIIVHLNKSVVINCTRPSNNTRTSITIGPGQVFYRTGDIIGDIRKAYCEINGTEWNKALKQVTEKLKEH
FNNKPIIFQPPSGGDLEITMHHFNCRGEFFYCNTTRLFNNTCIANGTIEGCNGNITLPCKIKQIINMWQGA
GQAMYAPPISGTINCVSNITGILLTRDGGATNNTNNETFRPGGGNIKDNWRNELYKYKVVQIEPLGVAPT
RAKRRVVEREKR

6.6.2 Human Antibody Fab Model Generation

Antibody Fab crystal structures used in the benchmark (Appendix - Table 6.1) and two additional antibodies (PDB ID: 2NY7 and 3HMX) were downloaded from the Protein Databank (PDB). Given the downstream application of our method to HIV glycoprotein gp120, primarily a human disease, and the availability of human sera microarray data, we chose to focus on human antibodies as potential antibody probes. We would like to note that this is a case-specific choice and it may be prudent to use an antibody probe from an organism for which immunogenicity is to be predicted. Human Fab models corresponding to appropriate heavy/light chain pairs were manually extracted from the PDB for each structure investigated. Execution of the patch prediction code generates and uses a Fab coordinate set corresponding to all model heavy atoms with occupancy greater than 0.5 and at most one atom position per unique atom identity.

6.6.3 Rosetta-Based Conformational Ensemble Modeling

Protein sequence for each RV144 component was threaded onto the crystal structure of gp120 (PDB ID: 3JWD) (*86*) using RosettaRemodel (*78*). Single positions were handled with fixed backbone design, any insertions and deletions were build *de novo* using local flexible backbone design functionality within RosettaRemodel. Conformational ensemble explored several regions, known to be flexible, such as V1/V2, V3, V4 and V5 variable loops, as well as the N and C termini. It is important to note that gD tag was modeled as a part of the N-termini for MN and CM244 components. Starting from the complete homology model, 5000

models were generated using RosettaRelax (*112*), from which the lowest-scoring 100 were selected to seed a new ensemble of 500 models, exploring the variability in the beta20/21, v4, v5 and N- and C-termini.

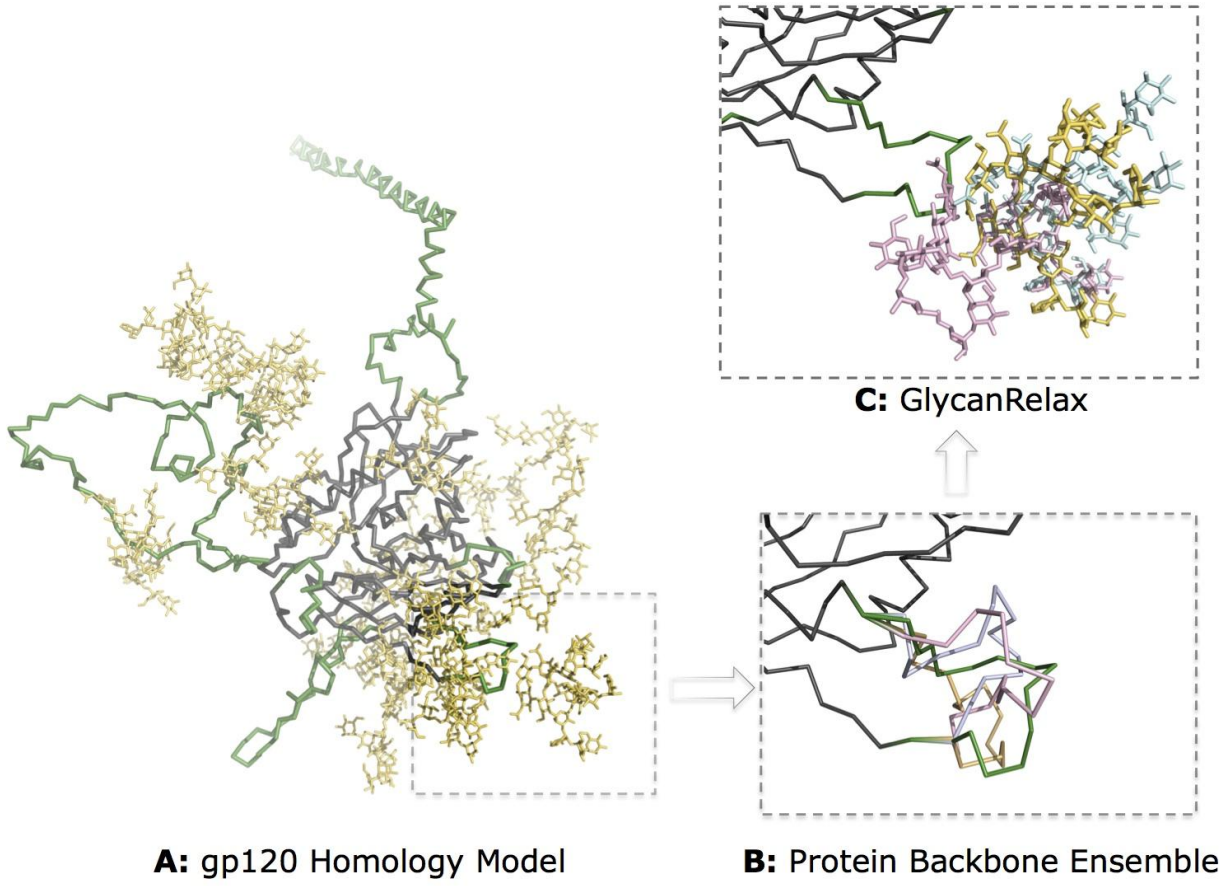


Figure 6.7: Rosetta Backbone Modeling and Glycan Relaxation. Regions of gp120 with known structural variability (green, panel A) are used to generate candidate loop conformations (panel B); the top 2% of gp120 conformations were used for subsequent glycan relaxation for each of the candidate loop conformations (panel C).

6.6.4 Glycan Relaxation

Each of the 500 models for a given component was decorated with glycans by aligning a Man8 N-glycan moiety on each asparagine within an appropriate N-X-(S/T). The Glycan Relax software used to generate models for EPIMAP prediction has been previously reported (*86*, *117*); it performs a two-stage manipulation of interatomic rotational dihedrals for glycan sugars and the underlying asparagine residue using a Monte Carlo procedure to ameliorate major steric clashes followed by cyclic coordinate descent (CCD) to minimize scoring function penalties associated with the Monte Carlo solution. Each model was used to seed two glycan minimization trajectories consisting of 10000 Monte Carlo and 10000 conformer descent (local optimization) trials. The resulting 1000 models were submitted for patch prediction.

6.6.5 Patch Prediction Software

The patch prediction code used to analyze conformational ensembles was coded as a unified executable file in Python according to the 2.X language version, using object-oriented class implementations of coordinate and coordinate set files to store protein databank models in indexable forms amenable to mathematical function calls. Care was given to render the underlying data structure classes as transparent as possible for future research.

Antibody positioning for consideration of patches is largely described in the Results section above; details of methodological interest but little relevance for interpretation are discussed here. Initialization of Fab models used for patch prediction requires spatial prepositioning for subsequent rotation and translation operations; the antibody's alignment axis is translated to (x,y,z) = (0,0,1) and the Fv is rotated into alignment with the (0,0,1) axis.

Non-CDR regions used for steric clash scoring are used to determine a minimal value of z minus Van der Waals radius for the Fab model, and the model is subsequently translated in the +z direction to ensure all volume occupied by its constituent atoms lies above the z=0 plane in the +z direction.

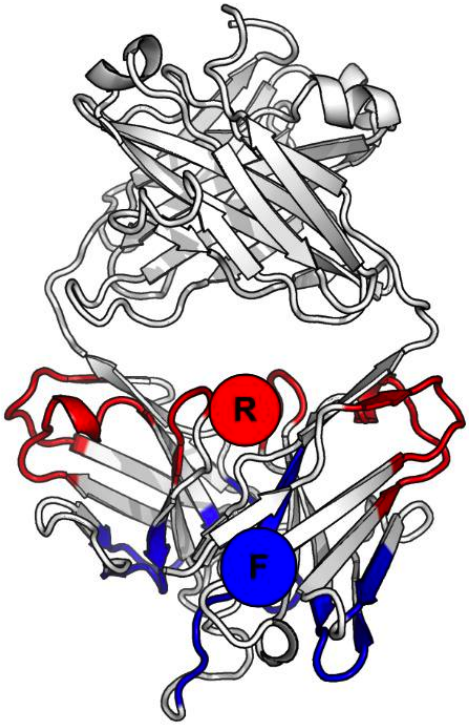


Figure 6.8: F_{AB} Model Probe Alignment Zones. Residues in the rear (R) and front (F) of the F_V are used to generate average positions; these form the vector, which is used to align the probe to orientation vectors during patch prediction (Fig. 6.3, center panels).

Alignment of Fab models to orientation vectors is determined by sequential rotation of an axis formed by two centers of mass corresponding to the binding surface and rear of the Fv; binary docking of the Fab model to the Patch identification is determined by a count of all heavy atoms (C, N, O, S) in the viral glycoprotein model within a 6Å interaction distance cutoff of a subset of Fab residues corresponding to CDR loops, again determined using the standard numbering system common to structures of human antibodies in the Protein Databank.

While the use of global variables is often considered to be bad programming practice, single-file executable focused on a limited subset of unique data structures can offer improved readability through judicious and limited global variable usage. Global variables of functional importance for EPIMAP are the dictionary of spatially binned glycoprotein atoms and floating point threshold parameters, such as interaction cutoff distances and patch count definitions, which might otherwise cause errors if left unset by scope.

6.6.6 Computing Hardware

Computation was performed at the Scripps Research Institute on an IAVI parallel computing cluster containing heterogeneous nodes:

Node Type	Cores/Node	Core Clock GHz	RAM/Node	# of Nodes
Silver	32	2.3	128 GB	28
Imperial	24	2.3	64 GB	10
Loma	12	3.5	72 GB	10

Processing for one executable, covering all patch predictions for one glycoprotein model, runs for 175 minutes on a 2011-edition iMac running Mac OS 10.7.5 with a 2.5 GHz Core i5 processor and 4 GB of 1333 MHz DDR3 RAM in the 2.7.1 version of Python.

6.7 Acknowledgements

EPIMAP has been produced by the joint efforts of Chris Carrico and Sergey Menis; Sergey ran the Rosetta-based modeling and managed cluster submissions, Chris developed the patch prediction software, and glycan relaxation and data analysis were joint projects. William Schief provided oversight and contributed to the overall software design and results analysis; Peter Gilbert provided the RV144 test case, the initial project specification and Raphael Gottardo with Allan DeCamp - the microarray data for experimental validation.

This dissertation chapter is in the process of adaptation into a manuscript for submission to *PLoS Computational Biology*, with co-first authorship for Chris Carrico and Sergey Menis, William R. Schief as corresponding author, Raphael Gottardo and Allan DeCamp and possibly Peter Gilbert as additional authors.

6.8 Appendix

Table 6.1: Benchmark Results for Antibody Fab Structures

RCSB Structure Identifier	Native Ab Probe	2NY7 Probe	2NY7-Native	3HMX Probe	3HMX-Native	3HMX-2NY7	CDRH3 Length (vs. 2NY7)
1ADQ	0.632	0.474	-0.158	0.579	-0.053	0.105	-4
1G9M	0.733	0.733	0.000	0.933	0.200	0.200	1
1I9R	0.750	0.700	-0.050	0.850	0.100	0.150	-9
1JPS	0.682	0.500	-0.182	0.591	-0.091	0.091	-10
2B4C	0.789	0.737	-0.053	0.684	-0.105	-0.053	4
2CMR	0.600	0.550	-0.050	0.600	0.000	0.050	-8
2DD8	0.636	0.591	-0.045	0.682	0.045	0.091	-9
2FJG	1.000	1.000	0.000	1.000	0.000	0.000	-7
2FJH	0.643	0.643	0.000	0.714	0.071	0.071	-6
2H9G	0.667	0.600	-0.067	0.800	0.133	0.200	-7
2QAD	0.565	0.652	0.087	0.739	0.174	0.087	3
2QQN	1.000	0.909	-0.091	1.000	0.000	0.091	-4
2UZI	0.632	0.579	-0.053	0.684	0.053	0.105	-12
2VXS	1.000	1.000	0.000	1.000	0.000	0.000	-9
2WUB	0.667	0.583	-0.083	0.583	-0.083	0.000	-7
2XQB	0.565	0.522	-0.043	0.565	0.000	0.043	-2
2XRA	0.550	0.450	-0.100	0.600	0.050	0.150	-5
3B2U	0.818	0.636	-0.182	0.682	-0.136	0.045	-7
3BDY	0.867	0.800	-0.067	0.733	-0.133	-0.067	-7
3BE1	0.737	0.684	-0.053	0.789	0.053	0.105	-7
3CSY	0.833	0.778	-0.056	0.778	-0.056	0.000	-11
3EOA	0.857	0.786	-0.071	0.786	-0.071	0.000	-6
3G6D	0.737	0.737	0.000	0.737	0.000	0.000	-5
3GBM	0.800	0.850	0.050	0.750	-0.050	-0.100	-6
3GBN	0.789	0.789	0.000	0.684	-0.105	-0.105	-6
3GRW	0.429	0.429	0.000	0.500	0.071	0.071	0
3H42	0.500	0.500	0.000	0.536	0.036	0.036	-4
3HI1	0.680	0.520	-0.160	0.480	-0.200	-0.040	-3
3HI6	0.632	0.684	0.053	0.632	0.000	-0.053	-7
3HMX	0.739	0.478	-0.261	0.739	0.000	0.261	-8
3IDX	0.684	0.684	0.000	0.684	0.000	0.000	1
3K2U	0.682	0.545	-0.136	0.727	0.045	0.182	-9
3L5X	0.700	0.800	0.100	0.800	0.100	0.000	-7

3L5Y	0.667	0.583	-0.083	0.583	-0.083	0.000	-7
3LEV	0.643	0.643	0.000	0.714	0.071	0.071	4
3LQA	0.722	0.611	-0.111	0.722	0.000	0.111	-2
3LZF	0.619	0.476	-0.143	0.667	0.048	0.190	-4
3MA9	0.538	0.462	-0.077	0.500	-0.038	0.038	-6
3N85	0.593	0.630	0.037	0.667	0.074	0.037	-10
3NFP	0.609	0.609	0.000	0.696	0.087	0.087	-11
3NGB	0.667	0.533	-0.133	0.733	0.067	0.200	-6
3NH7	0.480	0.440	-0.040	0.520	0.040	0.080	-7
3P0Y	0.765	0.647	-0.118	0.647	-0.118	0.000	-6
3PGF	0.583	0.417	-0.167	0.542	-0.042	0.125	-7
3Q1S	0.500	0.571	0.071	0.500	0.000	-0.071	-1
3R1G	0.727	0.636	-0.091	0.727	0.000	0.091	-8
3RU8	0.607	0.536	-0.071	0.607	0.000	0.071	0
3SDY	0.789	0.684	-0.105	0.684	-0.105	0.000	-5
3SE8	0.568	0.405	-0.162	0.459	-0.108	0.054	-4
3SE9	0.731	0.577	-0.154	0.615	-0.115	0.038	-4
3SKJ	0.609	0.522	-0.087	0.652	0.043	0.130	1
3SM5	0.714	0.762	0.048	0.762	0.048	0.000	-1
3SOB	0.714	0.524	-0.190	0.667	-0.048	0.143	-1
3T2N	0.600	0.700	0.100	0.650	0.050	-0.050	-14
3THM	0.647	0.588	-0.059	0.588	-0.059	0.000	-1
3TJE	0.588	0.588	0.000	0.706	0.118	0.118	-1
3TYG	0.889	0.889	0.000	0.889	0.000	0.000	1
3U2S	0.800	0.800	0.000	0.900	0.100	0.100	10
3U4E	0.909	0.909	0.000	0.818	-0.091	-0.091	10
3U7Y	0.694	0.500	-0.194	0.444	-0.250	-0.056	-2
3ZTN	0.650	0.650	0.000	0.700	0.050	0.050	2
4D9Q	0.714	0.714	0.000	0.714	0.000	0.000	-12
4DAG	0.478	0.522	0.043	0.522	0.043	0.000	-4
Average:	0.689	0.636	-0.053	0.686	-0.003	0.051	-4

CHAPTER 7: DISCUSSION AND CONCLUSIONS

In this manuscript, I present an application of epitope scaffolding methods to design prospective antigens for Hepatitis C (HCV). The protein scaffolds, displaying the HCV1 bNAb epitope, designed in the course of this work, mimic the antibody-bound conformation perfectly, verified by X-ray crystallography, and elicit antibodies toward the HCV1 epitope in BALB/c mice. This case study demonstrates for the first time that epitope scaffolds can elicit neutralizing antibodies in animals. The first generation scaffolds designed for the PG9 epitope are an important step towards the elicitation of neutralizing anti-HIV antibodies and development of robust methods necessary to deal with the multitude of recently described glycan-dependent HIV epitopes. To improve immunogenicity of HCV and any future scaffold immunogens, I developed a virus-like multimerization platform for prospective immunogens using an engineered domain of HIV gp120 (eOD) as a proxy for other antigens. The platform readily displays variants of eOD and we are working towards displaying HCV scaffolds on the same particle, to prove general applicability. We now have monomeric scaffolds, which elicit neutralizing antibodies and a multimeric platform suitable to boost their immunogenicity in vivo. The utility of these scaffolds and the platform will be evaluated experimentally. To aid with analysis of the vaccine trials, testing these scaffolds and other immunogens, we developed a computational method for identification of prospective antibody epitopes on flexible and glycosylated proteins. This information can be used to prefilter epitopes for sieve analyses, as applied to the RV144 trial, and guide immunogen design by identifying regions most frequently presented to the immune system. The methods available for epitope scaffolding and protein engineering, at least in the scope of the Rosetta protein-modeling suite, are robust and can be readily applied to design prospective immunogens for any available high-resolution crystallized target. The HCV scaffolds designed in Chapter 2 present a well-defined linear epitope recognized by two different bNAbs, their design process followed the standard epitope scaffolding procedures

based on an initial superposition match, which provides the best possible mimicry for the epitope in the desired conformation. In the hands of an experienced protein engineer, this process is rapid and prospective designs were generated in less than two months. However, it is important to note, only 50% of the designs expressed were folded proteins and the successful designs were variants of just a few base scaffolds, suggesting that the best strategy for obtaining soluble scaffolds is to start from a large panel of different proteins. While, a potential strategy to improve this statistic is to use protein scaffolds from thermophiles, which are generally more tolerant of protein modifications, a more important and interesting avenue is to explore why certain scaffolds of similar size can and cannot tolerate the introduction of the epitope sequence. Conceptually, this question is closely tied with in the progress of the protein structure prediction and design fields. The design stage of the current scaffolding protocols is primarily focused on the immediate area around the epitope; while, it is becoming quite clear that in order to reliably produce successful scaffolds, a protein designer must consider the global effects of such local sequence changes on the viability of the protein. Using a large number of base scaffolds may also be useful to produce well-folded and properly glycosylated molecules to present glycan-dependent epitopes. As presented in Chapter 3, mimicking the backbone of the glycosylation motif may not be sufficient to guarantee expected glycosylation and glycan processing. Hence, at the moment, having several scaffolds based on different proteins of different size/origin is the most conservative strategy for designing well-folded molecules.

Current scaffolding efforts have consistently delivered prospective immunogens, which accurately recapitulate the antibody-bound conformations of several protective antibodies. In the HCV1 scaffolding initiative, the scaffolds designed for one anti-HCV antibody bound several other antibodies, which were isolated independently from several mouse systems, suggesting that the epitopes presented are biologically important and exist in a productive conformation. From a protein design standpoint, these scaffolds are a success. This, however, does not guarantee that the monomeric scaffolds will be immunogenic and will

receive the same type of processing as a virus would, potentially dramatically reducing its chances for reeliciting the bNAbs of interest. To address this issue, we have designed multimerization platforms displaying a repetitive array of scaffolds on the surface, in a spherical arrangement, in order to mimic the size and shape of the virus. The ease of synthesis, production and purification of this multimeric platform should make it accessible for any antigen of similar size and we have developed additional strategies to allow for larger antigens to be displayed in a similar fashion, as well.

Finally, given the importance of glycosylation in antibody-epitope interactions for a number of bNAbs, the computational protocol for the identification of putative epitopes is especially interesting. With additional work, it should be possible to start modeling protein glycan interactions, which would allow for the identification of putative mixed glycan-protein epitopes, thereby, focusing the epitope scaffolding and complementary subunit vaccination strategies to the relevant regions. Computational identification of putative epitopes is also a powerful tool useful in the analysis of immunization trials. As show in the RV144 HIV trial, the computational methods can be used to narrow down the number of regions to be evaluated in a sieve analysis, increasing its statistical power. Similarly, the method can be applied to identify exposed epitopes on the antigen, prior to immunization, and attempt to cover up or delete unwanted regions in advance.

Using the above-described protein engineering and epitope scaffolding methods and the multimerization platform, we now have a flexible platform suitable to supply a varied panel of antigens for immunization trials. We are also well-equipped to analyze the resulting data filtering out non-essential epitopes and prescreening the potential antigen constructs for optimal presentation of the desired regions.

BIBLIOGRAPHY

1. S. A. Plotkin, Correlates of protection induced by vaccination. *Clinical and vaccine immunology : CVI* **17**, 1055 (Jul, 2010).

2. M. D. Stanley A. Plotkin, W. A. Orenstein, P. A. Offit, *Vaccines: Stanley A. Plotkin, Walter A. Orenstein and Paul A. Offit*. (SAUNDERS W B Company, 2008).

3. Hepatitis B virus: a comprehensive strategy for eliminating transmission in the United States through universal childhood vaccination. Recommendations of the Immunization Practices Advisory Committee (ACIP). *MMWR. Recommendations and reports : Morbidity and mortality weekly report. Recommendations and reports / Centers for Disease Control* **40**, 1 (Nov 22, 1991).

4. S. E. Goldstone, S. Vuocolo, A prophylactic quadrivalent vaccine for the prevention of infection and disease related to HPV-6, -11, -16 and -18. *Expert review of vaccines* **11**, 395 (Apr, 2012).

5. L. Kong *et al.*, Structural basis of hepatitis C virus neutralization by broadly neutralizing antibody HCV1. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 9499 (Jun 12, 2012).

6. X. Gao *et al.*, Prevalence and trend of hepatitis C virus infection among blood donors in Chinese mainland: a systematic review and meta-analysis. *BMC infectious diseases* **11**, 88 (2011).

7. K. Mohd Hanafiah, J. Groeger, A. D. Flaxman, S. T. Wiersma, Global epidemiology of hepatitis C virus infection: New estimates of age-specific antibody to HCV seroprevalence. *Hepatology* **57**, 1333 (Apr, 2013).

8. J. Halliday, P. Klenerman, E. Barnes, Vaccination for hepatitis C virus: closing in on an evasive target. *Expert review of vaccines* **10**, 659 (May, 2011).

9. X. Forns, J. Bukh, R. H. Purcell, The challenge of developing a vaccine against hepatitis C virus. *Journal of hepatology* **37**, 684 (Nov, 2002).

10. B. Bartosch, J. Dubuisson, F. L. Cosset, Infectious hepatitis C virus pseudo-particles containing functional E1-E2 envelope protein complexes. *The Journal of experimental medicine* **197**, 633 (Mar 3, 2003).

11. J. C. Meunier *et al.*, Vaccine-induced cross-genotype reactive neutralizing antibodies against hepatitis C virus. *The Journal of infectious diseases* **204**, 1186 (Oct 15, 2011).

12. P. Garrone *et al.*, A prime-boost strategy using virus-like particles pseudotyped for HCV proteins triggers broadly neutralizing antibodies in macaques. *Science translational medicine* **3**, 94ra71 (Aug 3, 2011).

13. M. Law *et al.*, Broadly neutralizing antibodies protect against hepatitis C virus quasispecies challenge. *Nature medicine* **14**, 25 (Jan, 2008).

14. P. Zhang *et al.*, Depletion of interfering antibodies in chronic hepatitis C patients and vaccinated chimpanzees reveals broad cross-genotype neutralizing activity. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 7537 (May 5, 2009).

15. "Progress report 2011: Global HIV/AIDS response" (WHO UNAIDS, 2011).

16. K. Allers *et al.*, Evidence for the cure of HIV infection by CCR5Delta32/Delta32 stem cell transplantation. *Blood* **117**, 2791 (Mar 10, 2011).

17. D. R. Burton *et al.*, A Blueprint for HIV Vaccine Discovery. *Cell host & microbe* **12**, 396 (Oct 18, 2012).

18. N. M. Flynn *et al.*, Placebo-controlled phase 3 trial of a recombinant glycoprotein 120 vaccine to prevent HIV-1 infection. *The Journal of infectious diseases* **191**, 654 (Mar 1, 2005).

19. P. Pitisuttithum *et al.*, Randomized, double-blind, placebo-controlled efficacy trial of a bivalent recombinant glycoprotein 120 HIV-1 vaccine among injection drug users in Bangkok, Thailand. *The Journal of infectious diseases* **194**, 1661 (Dec 15, 2006).

20. J. P. Julien *et al.*, Crystal Structure of a Soluble Cleaved HIV-1 Envelope Trimer. *Science*, (Oct 31, 2013).

21. S. Phogat, R. Wyatt, Rational modifications of HIV-1 envelope glycoproteins for immunogen design. *Current pharmaceutical design* **13**, 213 (2007).

22. L. Stamatatos, L. Morris, D. R. Burton, J. R. Mascola, Neutralizing antibodies generated during natural HIV-1 infection: good news for an HIV-1 vaccine? *Nature medicine* **15**, 866 (Aug, 2009).

23. I. Mikell *et al.*, Characteristics of the earliest cross-neutralizing antibody response to HIV-1. *PLoS pathogens* **7**, e1001251 (2011).

24. L. Kong *et al.*, Supersite of immune vulnerability on the glycosylated face of HIV-1 envelope glycoprotein gp120. *Nature structural & molecular biology* **20**, 796 (Jul, 2013).

25. J. Huang *et al.*, Broad and potent neutralization of HIV-1 by a gp41-specific human antibody. *Nature* **491**, 406 (Nov 15, 2012).

26. M. B. Zwick *et al.*, Broadly neutralizing antibodies targeted to the membrane-proximal external region of human immunodeficiency virus type 1 glycoprotein gp41. *Journal of virology* **75**, 10892 (Nov, 2001).

27. D. R. Burton *et al.*, Efficient neutralization of primary isolates of HIV-1 by a recombinant human monoclonal antibody. *Science* **266**, 1024 (Nov 11, 1994).

28. T. Muster *et al.*, A conserved neutralizing epitope on gp41 of human immunodeficiency virus type 1. *Journal of virology* **67**, 6642 (Nov, 1993).

29. X. Wu *et al.*, Rational design of envelope identifies broadly neutralizing human monoclonal antibodies to HIV-1. *Science* **329**, 856 (Aug 13, 2010).

30. X. Wu *et al.*, Focused evolution of HIV-1 neutralizing antibodies revealed by structures and deep sequencing. *Science* **333**, 1593 (Sep 16, 2011).

31. J. F. Scheid *et al.*, Sequence and structural convergence of broad and potent HIV antibodies that mimic CD4 binding. *Science* **333**, 1633 (Sep 16, 2011).

32. J. Jardine *et al.*, Rational HIV Immunogen Design to Target Specific Germline B Cell Receptors. *Science*, (Mar 29, 2013).

33. N. A. Doria-Rose *et al.*, HIV-1 neutralization coverage is improved by combining monoclonal antibodies that target independent epitopes. *Journal of virology* **86**, 3393 (Mar, 2012).

34. L. E. McCoy, R. A. Weiss, Neutralizing antibodies to HIV-1 induced by immunization. *The Journal of experimental medicine* **210**, 209 (Feb 11, 2013).

35. D. R. Burton, Antibodies, viruses and vaccines. *Nature reviews. Immunology* **2**, 706 (Sep, 2002).

36. T. Zhou *et al.*, Structural definition of a conserved neutralization epitope on HIV-1 gp120. *Nature* **445**, 732 (Feb 15, 2007).

37. R. M. Cardoso *et al.*, Broadly neutralizing anti-HIV antibody 4E10 recognizes a helical conformation of a highly conserved fusion-associated motif in gp41. *Immunity* **22**, 163 (Feb, 2005).

38. G. Ofek *et al.*, Structure and mechanistic analysis of the anti-human immunodeficiency virus type 1 antibody 2F5 in complex with its gp41 epitope. *Journal of virology* **78**, 10724 (Oct, 2004).

39. J. S. McLellan *et al.*, Structure of HIV-1 gp120 V1/V2 domain with broadly neutralizing antibody PG9. *Nature* **480**, 336 (Dec 15, 2011).

40. H. M. Berman *et al.*, The Protein Data Bank. *Nucleic acids research* **28**, 235 (Jan 1, 2000).

41. G. Ofek *et al.*, Elicitation of structure-specific antibodies by epitope scaffolds. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 17880 (Oct 19, 2010).

42. M. L. Azoitei *et al.*, Computational design of high-affinity epitope scaffolds by backbone grafting of a linear epitope. *Journal of molecular biology* **415**, 175 (Jan 6, 2012).

43. B. E. Correia *et al.*, Computational design of epitope-scaffolds allows induction of antibodies specific for a poorly immunogenic HIV vaccine epitope. *Structure* **18**, 1116 (Sep 8, 2010).

44. B. E. Correia *et al.*, Computational protein design using flexible backbone remodeling and resurfacing: case studies in structure-based antigen design. *Journal of molecular biology* **405**, 284 (Jan 7, 2011).

45. M. L. Azoitei *et al.*, Computation-guided backbone grafting of a discontinuous motif onto a protein scaffold. *Science* **334**, 373 (Oct 21, 2011).

46. J. S. McLellan *et al.*, Design and characterization of epitope-scaffold immunogens that present the motavizumab epitope from respiratory syncytial virus. *Journal of molecular biology* **409**, 853 (Jun 24, 2011).

47. L. Kong *et al.*, Structure of hepatitis C virus envelope glycoprotein E2 antigenic site 412 to 423 in complex with antibody AP33. *Journal of virology* **86**, 13085 (Dec, 2012).

48. T. J. Broering *et al.*, Identification and characterization of broadly neutralizing human monoclonal antibodies directed against the E2 envelope glycoprotein of hepatitis C virus. *Journal of virology* **83**, 12473 (Dec, 2009).

49. A. Owsianka *et al.*, Monoclonal antibody AP33 defines a broadly neutralizing epitope on the hepatitis C virus E2 envelope glycoprotein. *Journal of virology* **79**, 11095 (Sep, 2005).

50. A. W. Tarr *et al.*, Determination of the human antibody response to the epitope defined by the hepatitis C virus-neutralizing monoclonal antibody AP33. *The Journal of general virology* **88**, 2991 (Nov, 2007).

51. T. J. Morin *et al.*, Human monoclonal antibody HCV1 effectively prevents and treats HCV infection in chimpanzees. *PLoS pathogens* **8**, e1002895 (2012).

52. D. A. Calarese *et al.*, Antibody domain exchange is an immunological solution to carbohydrate cluster recognition. *Science* **300**, 2065 (Jun 27, 2003).

53. R. Pejchal *et al.*, A potent and broad neutralizing antibody recognizes and penetrates the HIV glycan shield. *Science* **334**, 1097 (Nov 25, 2011).

54. L. M. Walker *et al.*, Broad neutralization coverage of HIV by multiple highly potent antibodies. *Nature* **477**, 466 (Sep 22, 2011).

55. R. M. Zinkernagel, On natural and artificial vaccinations. *Annual review of immunology* **21**, 515 (2003).

56. F. Schödel, D. Peterson, J. Hughes, R. Wirtz, D. Milich, Hybrid hepatitis B virus core antigen as a vaccine carrier moiety: I. Presentation of foreign epitopes. *Journal of Biotechnology* **44**, 91 (1/26/, 1996).

57. E. V. Grgacic, D. A. Anderson, Virus-like particles: passport to immune recognition. *Methods* **40**, 60 (Sep, 2006).

58. T. Defrance, M. Taillardet, L. Genestier, T cell-independent B cell memory. *Current opinion in immunology* **23**, 330 (Jun, 2011).

59. L. X. Doan, M. Li, C. Chen, Q. Yao, Virus-like particles as HIV-1 vaccines. *Reviews in Medical Virology* **15**, 75 (2005).

60. E. T. Crooks *et al.*, A comparative immunogenicity study of HIV-1 virus-like particles bearing various forms of envelope proteins, particles bearing no envelope and soluble monomeric gp120. *Virology* **366**, 245 (Sep 30, 2007).

61. M. F. Bachmann, G. T. Jennings, Vaccine delivery: a matter of size, geometry, kinetics and molecular patterns. *Nature reviews. Immunology* **10**, 787 (Nov, 2010).

62. Y. El-Manzalawy, V. Honavar, Recent advances in B-cell epitope prediction methods. *Immunome research* **6 Suppl 2**, S2 (2010).

63. U. Kulkarni-Kale, S. Bhosle, A. S. Kolaskar, CEP: a conformational epitope prediction server. *Nucleic acids research* **33**, W168 (Jul 1, 2005).

64. P. Haste Andersen, M. Nielsen, O. Lund, Prediction of residues in discontinuous B-cell epitopes using protein 3D structures. *Protein science : a publication of the Protein Society* **15**, 2558 (Nov, 2006).

65. R. Rapberger, A. Lukas, B. Mayer, Identification of discontinuous antigenic determinants on proteins based on shape complementarities. *Journal of molecular recognition : JMR* **20**, 113 (Mar-Apr, 2007).

66. J. Ponomarenko *et al.*, ElliPro: a new structure-based tool for the prediction of antibody epitopes. *BMC bioinformatics* **9**, 514 (2008).

67. M. J. Sweredoski, P. Baldi, PEPITO: improved discontinuous B-cell epitope prediction using multiple distance thresholds and half sphere exposure. *Bioinformatics* **24**, 1459 (Jun 15, 2008).

68. X. Zhang, W. Meining, M. Fischer, A. Bacher, R. Ladenstein, X-ray structure analysis and crystallographic refinement of lumazine synthase from the hyperthermophile Aquifex aeolicus at 1.6 A resolution: determinants of thermostability revealed from structural comparisons. *Journal of molecular biology* **306**, 1099 (Mar 9, 2001).

69. M. Rolland *et al.*, Increased HIV-1 vaccine efficacy against viruses with genetic signatures in Env V2. *Nature* **490**, 417 (Oct 18, 2012).

70. B. N. Fields, D. M. Knipe, P. M. Howley, *Fields' Virology*. (Wolters Kluwer Health/Lippincott Williams & Wilkins, 2007).

71. B. D. Lindenbach *et al.*, Complete replication of hepatitis C virus in cell culture. *Science* **309**, 623 (Jul 22, 2005).

72. M. Dorner *et al.*, Completion of the entire hepatitis C virus life cycle in genetically humanized mice. *Nature*, (Jul 31, 2013).

73. A. Ploss *et al.*, Human occludin is a hepatitis C virus entry factor required for infection of mouse cells. *Nature* **457**, 882 (Feb 12, 2009).

74. W. O. Osburn *et al.*, Spontaneous control of primary hepatitis C virus infection and immunity against persistent reinfection. *Gastroenterology* **138**, 315 (Jan, 2010).

75. Z. Stamataki, S. Coates, S. Abrignani, M. Houghton, J. A. McKeating, Immunization of human volunteers with hepatitis C virus envelope glycoproteins elicits antibodies

that cross-neutralize heterologous virus strains. *The Journal of infectious diseases* **204**, 811 (Sep 1, 2011).

76. K. Yusim *et al.*, Genotype 1 and global hepatitis C T-cell vaccines designed to optimize coverage of genetic diversity. *The Journal of general virology* **91**, 1194 (May, 2010).

77. Schrodinger, LLC. (2010).

78. P. S. Huang *et al.*, RosettaRemodel: a generalized framework for flexible backbone protein design. *PloS one* **6**, e24109 (2011).

79. M. S. Weiss, R. Hilgenfeld, On the use of the merging R factor as a quality indicator for X-ray data. *Journal of Applied Crystallography* **30**, 203 (1997).

80. Z. Otwinowski, W. Minor, in *Methods in enzymology,* Charles W. Carter, Jr., Ed. (Academic Press, 1997), vol. Volume 276, pp. 307-326.

81. P. D. Adams *et al.*, PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta crystallographica. Section D, Biological crystallography* **66**, 213 (Feb, 2010).

82. R. F. Clayton *et al.*, Analysis of Antigenicity and Topology of E2 Glycoprotein Present on Recombinant Hepatitis C Virus-Like Particles. *Journal of virology* **76**, 7672 (August 1, 2002, 2002).

83. J. R. Mascola, D. C. Montefiori, HIV-1: nature's master of disguise. *Nature medicine* **9**, 393 (Apr, 2003).

84. P. R. Clapham, S. Lu, Vaccinology: precisely tuned antibodies nab HIV. *Nature* **477**, 416 (Sep 22, 2011).

85. A. I. Rico, M. Garcia-Ovalle, J. Mingorance, M. Vicente, Role of two essential domains of Escherichia coli FtsA in localization and progression of the division ring. *Molecular microbiology* **53**, 1359 (Sep, 2004).

86. M. Pancera *et al.*, Structure of HIV-1 gp120 with gp41-interactive region reveals layered envelope architecture and basis of conformational mobility. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 1166 (Jan 19, 2010).

87. D. R. Burton, P. Poignard, R. L. Stanfield, I. A. Wilson, Broadly neutralizing antibodies present new prospects to counter highly antigenically diverse viruses. *Science* **337**, 183 (Jul 13, 2012).

88. W. R. Schief, Y. E. Ban, L. Stamatatos, Challenges for structure-based HIV vaccine design. *Current opinion in HIV and AIDS* **4**, 431 (Sep, 2009).

89. D. C. Ekiert *et al.*, Antibody recognition of a highly conserved influenza virus epitope. *Science* **324**, 246 (Apr 10, 2009).

90. J. P. Julien, P. S. Lee, I. A. Wilson, Structural insights into key sites of vulnerability on HIV-1 Env and influenza HA. *Immunological reviews* **250**, 180 (Nov, 2012).

91. T. Zhou *et al.*, Structural basis for broad and potent neutralization of HIV-1 by antibody VRC01. *Science* **329**, 811 (Aug 13, 2010).

92. R. Diskin *et al.*, Increasing the potency and breadth of an HIV antibody by using structure-based rational design. *Science* **334**, 1289 (Dec 2, 2011).

93. A. P. West, Jr., R. Diskin, M. C. Nussenzweig, P. J. Bjorkman, Structural basis for germ-line gene usage of a potent class of antibodies targeting the CD4-binding site of HIV-1 gp120. *Proceedings of the National Academy of Sciences of the United States of America* **109**, E2083 (Jul 24, 2012).

94. R. Arnaout *et al.*, High-resolution description of antibody heavy-chain repertoires in humans. *PloS one* **6**, e22365 (2011).

95. A. Leaver-Fay *et al.*, ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Methods in enzymology* **487**, 545 (2011).

96. G. Chao *et al.*, Isolating and engineering human antibodies using yeast surface display. *Nature protocols* **1**, 755 (2006).

97. C. Genomes Project *et al.*, An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56 (Nov 1, 2012).

98. E. Krissinel, K. Henrick, Inference of macromolecular assemblies from crystalline state. *Journal of molecular biology* **372**, 774 (Sep 21, 2007).

99. E. F. Pettersen *et al.*, UCSF Chimera--a visualization system for exploratory research and analysis. *Journal of computational chemistry* **25**, 1605 (Oct, 2004).

100. T. Ota *et al.*, Anti-HIV B Cell lines as candidate vaccine biosensors. *Journal of immunology (Baltimore, Md. : 1950)* **189**, 4816 (Nov 15, 2012).

101. S. Hoot *et al.*, Recombinant HIV envelope proteins fail to engage germline versions of anti-CD4bs bNAbs. *PLoS pathogens* **9**, e1003106 (Jan, 2013).

102. C. Sundling *et al.*, Soluble HIV-1 Env trimers in adjuvant elicit potent and diverse functional B cell responses in primates. *The Journal of experimental medicine* **207**, 2003 (Aug 30, 2010).

103. K. L. Knight, C. R. Winstead, Generation of antibody diversity in rabbits. *Current opinion in immunology* **9**, 228 (Apr, 1997).

104. B. F. Haynes, G. Kelsoe, S. C. Harrison, T. B. Kepler, B-cell-lineage immunogen design in vaccine development with HIV-1 as a case study. *Nature biotechnology* **30**, 423 (May, 2012).

105. A. T. McGuire *et al.*, Engineering HIV envelope protein to activate germline B cell receptors of broadly neutralizing anti-CD4 binding site antibodies. *The Journal of experimental medicine* **210**, 655 (Apr 8, 2013).

106. B. Yao, D. Zheng, S. Liang, C. Zhang, Conformational B-cell epitope prediction on antigen protein structures: a review of current algorithms and comparison with common binding site prediction methods. *PloS one* **8**, e62249 (2013).

107. R. Das, D. Baker, Macromolecular modeling with rosetta. *Annual review of biochemistry* **77**, 363 (2008).

108. R. R. Garrity *et al.*, Refocusing neutralizing antibody response by targeted dampening of an immunodominant epitope. *Journal of immunology (Baltimore, Md. : 1950)* **159**, 279 (Jul 1, 1997).

109. S. Sampath *et al.*, Glycan masking of Plasmodium vivax Duffy Binding Protein for probing protein binding function and vaccine development. *PLoS pathogens* **9**, e1003420 (Jun, 2013).

110. M. E. Ackerman *et al.*, Natural variation in Fc glycosylation of HIV-specific antibodies impacts antiviral activity. *The Journal of clinical investigation* **123**, 2183 (May 1, 2013).

111. C. K. Leonard *et al.*, Assignment of intrachain disulfide bonds and characterization of potential glycosylation sites of the type 1 recombinant human immunodeficiency virus envelope glycoprotein (gp120) expressed in Chinese hamster ovary cells. *The Journal of biological chemistry* **265**, 10373 (Jun 25, 1990).

112. C. Wang, P. Bradley, D. Baker, Protein-protein docking with backbone flexibility. *Journal of molecular biology* **373**, 503 (Oct 19, 2007).

113. P. D. Kwong *et al.*, Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature* **393**, 648 (Jun 18, 1998).

114. P. Liu *et al.*, Infectious Virion Capture by HIV-1 gp120-Specific IgG from RV144 Vaccinees. *Journal of virology* **87**, 7828 (Jul, 2013).

115. S. M. Alam *et al.*, Antigenicity and immunogenicity of RV144 vaccine AIDSVAX clade E envelope immunogen is enhanced by a gp120 N-terminal deletion. *Journal of virology* **87**, 1554 (Feb, 2013).

116. P. W. Berman, Development of bivalent rgp120 vaccines to prevent HIV type 1 infection. *AIDS research and human retroviruses* **14 Suppl 3**, S277 (Oct, 1998).

117. J. M. Binley *et al.*, Role of complex carbohydrates in human immunodeficiency virus type 1 infection and resistance to antibody neutralization. *Journal of virology* **84**, 5637 (Jun, 2010).