Understanding and Manipulating Anti-HIV Antibody Responses via Structural Methods

Christopher T Carrico

A dissertation

Submitted in partial fulfillment of the

Requirements for the degree of

Doctor of Philosophy

University of Washington

2013

Reading Committee:

Roland Strong, Chair

Phil Bradley

Joe Smith

Program Authorized to Offer Degree:

Biochemistry

University of Washington

Abstract

Understanding and Manipulating Anti-HIV Antibody Responses via Structural Methods

Christopher T Carrico

Chair of the Supervisory Committee:

Roland Strong, Full Member

Fred Hutchinson Cancer Research Center

Anti-HIV antibody responses offer one of very few potential routes towards a protective HIV vaccine; structural biochemistry methods such as crystallography and computational modeling can provide key insights for the design of candidate vaccine components and enhance the evaluation of complex experimental outcomes. Accordingly, described herein are four projects investigating antibody interactions with HIV: (1) the design of scaffolds incorporating the V1/V2 loop of HIV gp120 for the determination of its structure as bound to the broadly neutralizing anti-HIV antibody PG9, (2) the determination of antibody exposure for the protein components used in the RV144 vaccine trial, (3) crystallographic analysis of designed proteins presenting the b12 antibody's epitope as transplanted from HIV's gp120 envelope glycoprotein, and (4) biophysical characterization of b12's interactions with gp120 and development from its germline precursor antibody sequence using a panel of point mutants. These efforts, representing work on multiple aspects of potentially key importance in shifting basic science research on HIV vaccinology into an translational and ultimately clinical phase, are representative of current work in B cell vaccine research for both the HIV field and for work on other pathogens of current interest.

For Rhiannon, for pushing me back into all this.

Contents:

Chapter 1: An Overview of Antibodies, Neutralization, B Cells and HIV Vaccination Responses

Why Focus on Antibodies?

Of all the medical interventions in the modern repertoire, few are as effective or as cost-effective as the prevention of infectious disease by vaccination. Pathogens responsible for millions of deaths have been eradicated or contained; technological refinement in vaccine design has broadened the field from crude initial attempts with heterologous live pathogens such as cowpox to recombinant subunit vaccines with near-perfect safety and efficacy. However, progress in vaccine design and deployment has not progressed uniformly across all fronts; certain key pathogens such as HIV-1[1] and malaria[2] have resisted decades of intensive effort to develop even modestly effective vaccines, while others such as influenza have settled into a cyclic battle of escape and protection wherein current best practice can provide protection only imperfectly and through continual and expensive effort.

The key to vaccine efficacy is successful presentation of antigenic epitopes from a target pathogen to the adaptive immune system, with the subsequent elicitation of a protective response against the pathogen. Such protective responses are mediated by lymphocytes termed T and B cells; T cells may be usefully subdivided into both local effector cells which directly combat pathogens and immunoregulatory cells which help marshal and limit immune responses. B cells, in contrast, serve to produce proteins - known as antibodies - specific for fixed epitopes displayed either directly by a pathogen of interest or by cells it has infected.

The immune response against a particular pathogen, and the elements thereof responsible for protection after successful vaccination, typically consists of both T and B cell responses. Much current research focuses on both the two components separately and the interplay on various levels between them; the work presented here, however, will focus almost exclusively on the B cell response, with limited discussion of assistance a successful B cell response can provide to other immune functions. The soluble antibodies produced by B cells are particularly amenable to biochemical study, unlike the more complicated cell biology intrinsic to effector T cell function; B cell-derived antibodies can also offer key insights into pathogen structure and function which are not observable through even the most direct T cell interactions.

Basic Aspects of Antibody Structure and Function

Antibodies - also termed immunoglobulins (Igs), and in their initial stages B-cell receptors (BCRs) - are secreted proteins produced by clonal B cell lineages and specific for a particular target antigen. An antibody is a polyvalent arrangement of heterodimeric subunits (Fig. 1.2A), each of which consists of a heavy and a light chain; both the heavy and light chains possess two N-terminal Ig-fold domains which interact to form an antigen-binding region (Fab), the foremost of which pair to form the variable fragment (Fv) which bears the antigen-contacting surface. Additionally, the heavy chain possesses two or more constant domains which participate in inter-subunit oligomerization.

To achieve a high level of specificity and affinity towards a broad range of potential target antigens, antibodies must exhibit significant structural variability; antibodies possess heterogeneous sequences, particularly in the complementarity-determining regions (CDRs) which form the antigen-binding surface of the Fv (Fig 1.2B), and exhibit a considerable range of lengths, conformations, charges, and post-translational modifications.

Antibodies initially occur as membrane-bound B cell surface receptors (BCRs); following B cell activation and proliferation via successful binding of BCR to antigen (Fig. 1.2C), the receptor-bearing B cell may begin production of alternate classes of antibody with different oligomeric states and downstream effector functions (Fig. 1.2D). BCRs are initially expressed as IgM, which occurs as a monomer on the B cell surface or as a highly avid secreted pentamer; IgD expression begins as a mature B cell first exits the bone marrow (as discussed below). IgA, IgE and IgG are the product of terminally differentiated effector B cell types; IgA is primarily involved in mucosal immunity, IgE in the activation of other specialized non-B lymphocytes, and IgG in the canonically understood antibody protection on which this work focuses. IgG may be further subdivided into four subtypes with different effector functions; their mode of antigen recognition is, however, identical.

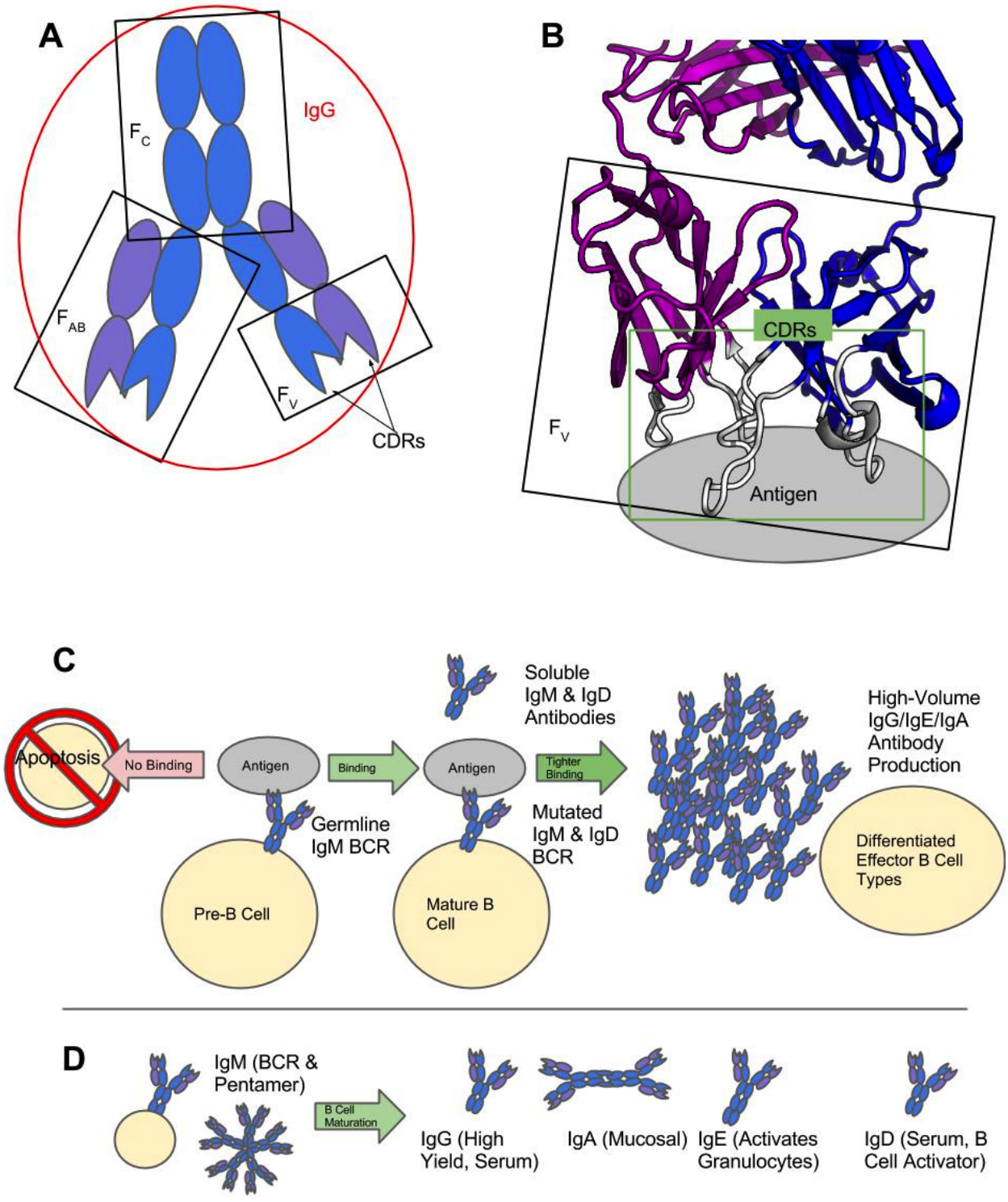


Figure 1.1: Antibody Structure and Classes. Antibodies consist of multiple heavy chains (blue) partnered with light chains (purple); as typically understood, two heavy chains and their respective light chains pair together to form an immunoglobulin, with two N-terminal F_{AB} and one C-terminal F_C regions **(A)**. The variable fragment (F_V) within the F_{AB} contains the complementarity-determining regions (CDRs) responsible for antigen binding **(A,B)**. Soluble antibody production by B cells requires proliferative simulation via membrane-bound BCRs by antigen, and high-volume production is performed by

differentiated B cells displaying little or no surface BCR **(C).** A diverse array of antibody classes are produced following B cell maturation, with different localizations and immunological roles **(D).**

Somatic Hypermutation and B Cell Biology

Antibody sequences' considerable genetic variability is intrinsic to their function as specific, adaptive mediators of immune function against particular pathogens. This variability is achieved through two key stages early in the development of the B cell bearing a particular antibody sequence; thus, some understanding of B cell biology assists in understanding the development of antibody specificity and affinity.

B cells originate in the bone marrow from hematopoietic progenitors (Fig. 1.3), undergoing an initial process of BCR generation prior to a selection process based on reactivity to so-called "self" antigens. Selection of pre-B cells on the basis of self antigen reactivity has the important function of deleting autoreactive B cells capable of creating pathological autoimmune antibodies[3]. Self reactivity is blockaded both by active deletion of B cells with BCRs showing high affinity to cell-surface self antigens, and by a process termed anergy by which B cells demonstrating weaker affinity persist but do not successfully enter the subsequent stages of B cell development.

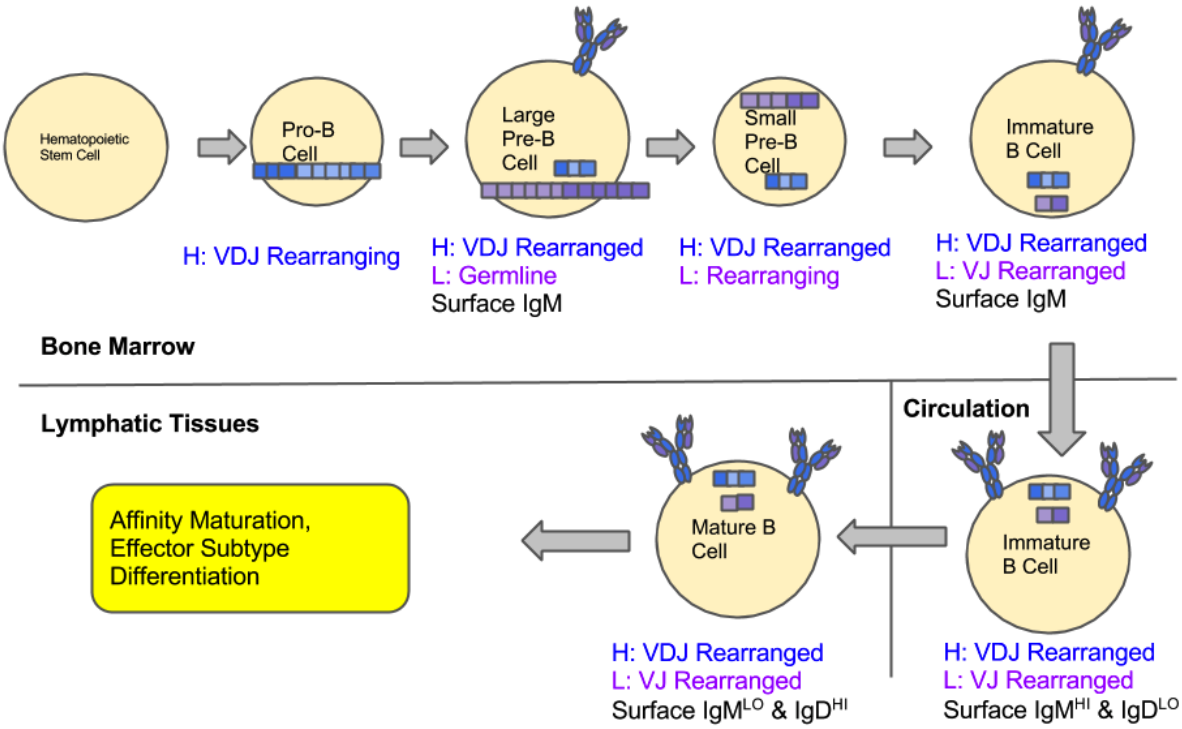


Figure 1.2: B Cell Early Development and V(D)J Rearrangement Pathway. The heavy and light chains are rearranged sequentially prior to B cells entering the lymphatic tissue for subsequent hypermutation.

The initial BCR sequence presented on a pre-B cell is the product of a genetic shuffling operation by which multiple gene segments - located adjacently within the same gene locus - are recombined to excise all but one of each segment type from the final protein sequence (Fig. 1.4A). These subsections of the BCR are termed the V and J genes for the N and C termini of the protein sequence, respectively. Additionally, the BCR heavy chain (but not the light chain) possesses a D gene segment intercalated between its V and J segments. BCR heavy chain recombination proceeds to completion prior to light chain recombination; in addition to the shuffling of multiple sequences to achieve combinatorial diversity, random nucleotide addition is performed at the gene segment junctions, resulting in so-called junctional diversity.

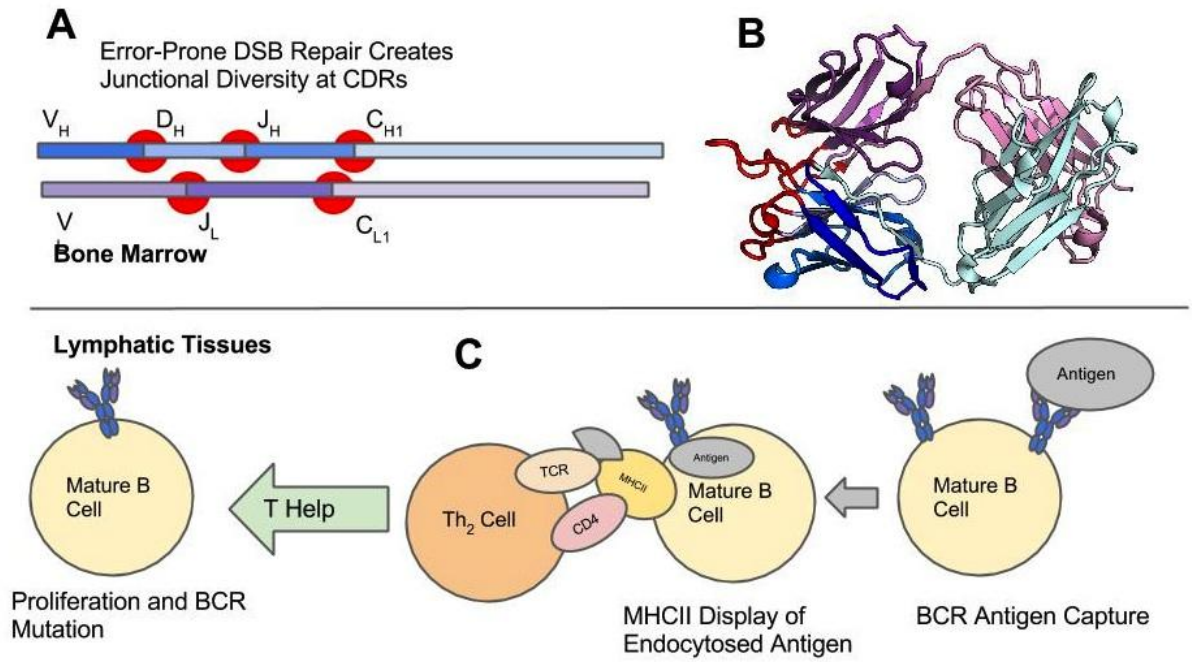


Figure 1.3: Junctional Diversity is Followed by Competitive, Mutating Proliferation. Initial V(D)J recombination in the bone marrow **(A)** creates mismatches between DNA segments which are resolved in an error-prone fashion by DNA double-strand-break repair, leading to sequence diversity at the V(D)J junctions; these form the antigen-binding CDRs, seen here in red on the 3RU8 crystal structure of the b12 F_{AB}. Subsequently, mature B cells compete for antigen in the lymphatic tissue, acquiring the helper T cell stimulation they require for proliferation and the accumulation of further affinity-enhancing mutations **(C).**

Following V(D)J recombination completion, a pre-B cell must be activated by exposure to antigen to achieve proliferation and improved affinity for that antigen; pre-B cells without such antigenic exposure apoptose. The canonical mode of naive B cell activation after a newly mature B cell reaches the lymphatic tissue requires class II T helper cells (Th2) to recognize proteolyzed segments of antigen displayed on pre-B-surface MHC class II complexes; this antigen proteolysis is accomplished following B cell endocytosis of BCR-bound antigen, necessitating successful BCR:antigen binding for the B cell to receive so-called T help (Fig. 1.4B). As a B cell proliferates following T help, cytosine deaminases create further mutations in the cell's BCR sequence; the resulting daughter cells may, if the mutations created are advantageous for antigen binding or BCR expression, out-compete their progenitor for antigen and thus for T help. In this manner, cycles of proliferation and T help continue until sufficient affinity has been reached to allow further differentiation of the resultant B cells into effector-related subtypes. The iterative process of mutational accumulation and affinity maturation is termed somatic hypermutation, as the pre-B cell specifically and densely mutates the BCR-containing somatic gene.

Antibodies as typically encountered in the clinical setting are most efficiently produced by plasma cells[4], a fully differentiated endpoint of the B cell maturation pathway described above, which persist for weeks or months and do not divide; the longest-lived plasma cells depend upon a steady supply of antigen to continue antibody production, and in the absence of such antigen exposure plasma cell antibody output declines comparatively rapidly. In the case of a vaccine intended to provide protection over a period of years or decades without antigenic exposure, other B cell subtypes - such as memory B cells - reenter the proliferative cycle, ensuring a renewed supply of soluble antibodies specific to an invading pathogen in addition to potentially adding new mutations to the clonal repertoire.

Neutralization vs. Other Antibody Effects

Antibodies provide defense against pathogens through multiple mechanisms, the majority of which require other components of the immune system[5]. These may be broadly subdivided into cell-independent modes, such as neutralization, and the activation of effector cell types such as macrophages, natural killer cells, mast cells and granulocytes. Binding of an antibody's Fc region by a variety of specialized receptors (FcRs) mediates almost all of these defensive modes; Fc receptors exhibiting

variable specificities for different antibody classes are hallmarks of the myeloid lineages responsible for pathogen phagocytosis, chemokine responses, and the clearance of infected cells by the immune response. Additionally, the so-called complement system - consisting of a cascade of enzymes recruited via Fc binding - serves an important, antibody-dependent role in the recruitment and optimization of effector cell responses, as well as producing direct defensive effects such as cell lysis of targeted pathogens.

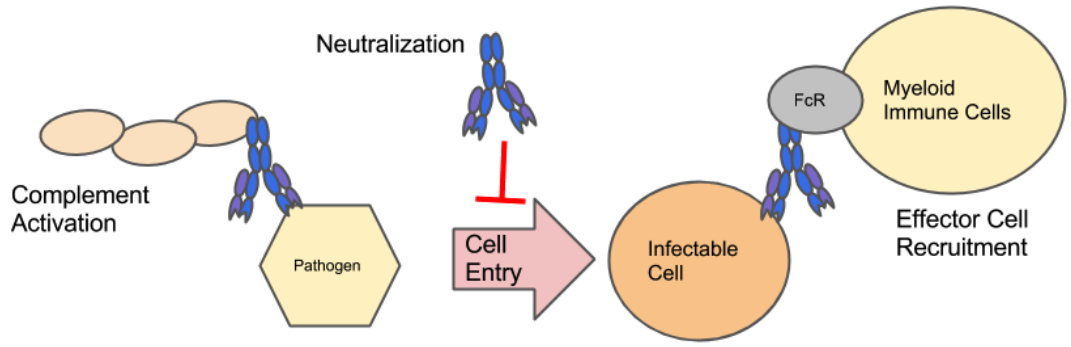


Figure 1.4: Protective Antibody Functions. In addition to activating the complement cascade and recruiting effector cell types via F_C binding, antibodies can directly neutralize pathogens, often by preventing cell-type-specific interactions necessary for infection.

An exception to the general Fc-based paradigm of antibody protection is antibody neutralization, in which binding to particular, functionally important epitopes on key pathogen components allows the inactivation or impairment of an entire pathogen without an explicit requirement for any Fc-mediated responses. Antibody neutralization often targets cell entry proteins on the outer membrane of a particular pathogen; the pathogen may be cleared by canonical Fc-dependent effector cell mechanisms in the usual fashion following such inactivation, whereas non-neutralizing antibody binding to the same protein might allow successful cell entry prior to effector cell arrival or complement-induced destruction of the target. As such, antibody neutralization is of particular interest for the design of vaccines against pathogens such as malaria, which exhibits a very short exposure time in the blood, or HIV-1, which can establish an essentially incurable infection based on a single successful invasion of a CD4+ lymphocyte[6].

Antibodies And HIV-1

The human immunodeficiency virus HIV-1 (HIV) has, in over three decades, grown into a global epidemic infecting tens of millions based largely on the reliable failure of the human immune system to mount a protective response capable of clearing or controlling the progression of its infection. Potentially protective responses in infected individuals have been observed, however, which - though insufficient to clear the virus following infection - would be sufficient to protect an uninfected individual from initial infection[7] with an efficacy in line with available vaccines for other major pathogenic viruses such as Hepatitis B. The two most promising immune response components which form the basis for current vaccine attempts are the control of replicating HIV by CD8-positive T cells, and the production of broadly neutralizing antibodies - so termed because of their capability to both bind and neutralize significant fractions of heterologous viral sequences.

Broadly neutralizing antibody (BNAb) responses against HIV demonstrate several atypicalities from responses against the acute-phase infections with which the human immune system typically contends. First, they develop later in infection; initial infection with HIV results in an acute-phase, highly polyclonal antibody response which generally fails to directly neutralize the virus[8], while neutralization develops only after several weeks and neutralization against broad subsets of viral sequence space only after years if at all[9]. Second, the development of antibody breadth is in itself unusual, as acute-phase responses to the majority of pathogenic human viruses have narrow specificity; some anti-HIV BNAbs of note have even been reported to be autoreactive[10], particularly those discovered in the 1990s and early 2000s, though the majority presently isolated are not. Some researchers have speculated that the immune dysregulation commonly seen in cases of chronic HIV infection is, in fact, necessary for the development of BNAbs; certainly, B cell dysregulation is known to occur in chronic HIV infection, and thus correlates with BNAb development.

One subfield of considerable current interest in HIV-related antibody research is the investigation of non-neutralizing activities of antibodies - including BNAbs - involved in protection against the virus in viral challenge experiments and recent vaccine trials. In particular, the identification of antibodies against a particular region of the HIV surface glycoprotein gp120 as a correlate of protection in the RV144 trial[11] implies that even non-neutralizing antibodies may offer some protection. Additionally, Fc-dependent antibody effects such as the induction of improved cellular cytotoxicity via complement and immune cell

recruitment have been shown to be important in achieving highly potent protection by BNAbs in an experimental animal model[12].

Current Frontiers in Antibody-Focused HIV Vaccine Design

Given the immense human and social costs associated with HIV, and the difficulties seen to date in controlling its spread, a vaccine of even modest protective efficacy is seen as a highly desirable tool for improving health worldwide. For the purposes of this dissertation, we will disregard the exciting progress being made in T cell-focused HIV vaccine research. Attempts towards B cell-focused vaccine candidates are currently grappling with two main questions, which are to some degree interrelated but often serve as primary foci for groups in the field. First, given that the bulk of clinically observed anti-HIV responses would fail to protect a vaccinee from infection, what are the immunologically accessible binding modes for potentially protective antibodies? Next, given that multiple avenues of attack have been found by successful BNAbs isolated from clinical patients, how can they be re-elicited by HIV-derived immunogens?

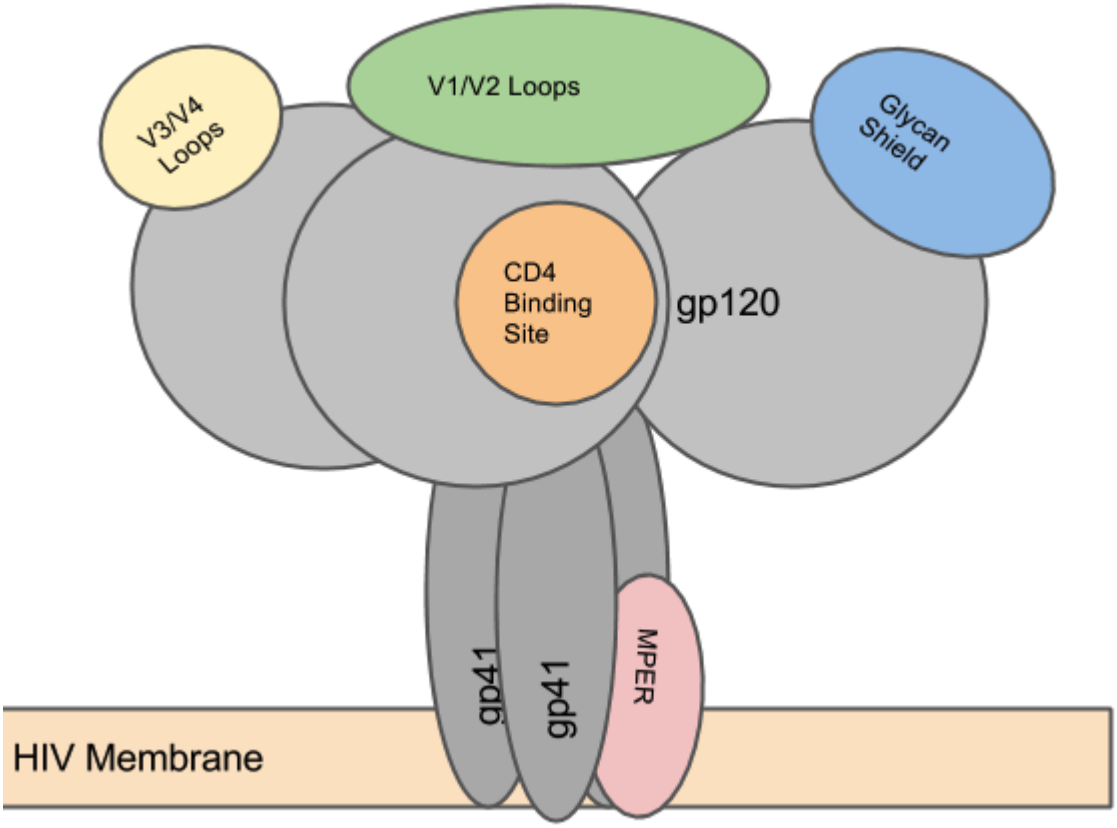


Figure 1.5: Schematic Representation of the HIV *Env* Trimer with BNAb Epitopes. See also Figure 3.1 in Chapter 3.

Antibody	Year of Discovery	Target Epitope
b12	1992[13][14]	CD4bs
2F5	1992[15][16]	gp41 MPER
4E10	2001[17]	gp41 MPER
447-52D	1993[18]	V3 Loop
2g12	1994[19]	Glycan Shield
VRC01	2010[20]	CD4bs
PG9	2009[21]	V1/V2 Loops

Table 1.1: Some Notable Anti-HIV Broadly Neutralizing Antibodies.

Since the discovery and isolation of 2F5 and b12 in 1992, dozens of antibodies against the two glycoprotein gene products of the HIV *Env* gene have been discovered; while initially discovered antibodies either potently or broadly neutralized viral primary isolates, refinements in single-B-cell

analysis recently enabled the identification of several clonal lineages of BNAbs targeting a variety of vulnerable epitopes, particularly on gp120. As the number and diversity of BNAbs identified has increased, the number of distinct targets used on the *Env* products' surface has leveled off; current data suggests that the sites of antigenic vulnerability have been mapped, but more potent and more broadly neutralizing antibodies continue to be discovered, implying that the upper limit of effectiveness against these few select targets has not been reached. Current benchmarks for antibody efficacy and breadth include the VRC01 family of antibodies against the CD4 binding site on gp120[22], PG9/16 against conformationally dependent epitopes on the gp120 V1/V2 loops[23], and the 10E8 antibody against the membrane-proximal region of gp41[24].

In spite of the newfound plethora of BNAb responses against HIV, vaccine-mediated elicitation of antibodies comparable to those recently identified has not been observed. As mentioned previously, acute infection with HIV succeeds in producing ample titers of antibodies, but these generally fail to neutralize even virus closely related to the infecting strain. If BNAb development during the course of HIV infection depends on the presentation of a multitude of related sequences which select for B cell responses capable of binding broad subsets of the available viral sequence space, then current methods of vaccine preparation are remarkably ill-equipped to recapitulate such antigenic exposure due to both insufficient duration of vaccination protocols and an emphasis on the preparation of biochemically homogeneous antigens rather than broadly heterogeneous antigen populations.

Current work in B cell vaccine research tends towards optimization and engineering of antigens, on the one hand, and refinement of sequential antigen presentation on the other with the shared goal of increasing the immune response against BNAb targets at the expense of *Env* epitopes which normally dominate the anti-HIV antibody response. Compelling evidence exists that rational selection or design of an optimized antigen can attenuate responses against undesired epitopes (see refs. [Depends on final merged list] and Appendix 1), whether by the more moderate technique of resurfacing an antigen of interest or by the more drastic step of transplanting a structurally intact B cell epitope onto a heterologous scaffolding protein.

Attempts to refine the administration schedules for vaccine candidates to foster protective immune responses are, similarly, ongoing; one concept of particular interest is the "prime-boost" strategy

for sequential presentation of related but distinct forms of a particular antigen. This strategy, often used with persistent cDNA priming followed by recombinant protein boosting, can both increase antibody titers and steer antibody responses towards desired epitopes; there thus exists considerable synergy between prime-boost vaccination and the use of engineered vaccine components.

Given the multivariate complexity of current proposed HIV vaccines, structural information regarding antibody interactions with HIV provides a key advantage in the selection and engineering of optimal antigens. Such structural information is, in the form of x-ray crystal structures of antibody:antigen complexes, being generated on an ongoing basis by consortia of research groups. However, as discovery of these desirable endpoints of B cell development in isolation has thus far been insufficient to engender their elicitation by vaccines, work in the field is expanding out into deeper analyses of the mechanisms responsible for antibody-mediated protection against HIV and the ontogenies of particularly successful BNAbs.

Work Presented In This Dissertation

Over the past five years, I have had the opportunity to work on multiple phases of the HIV B cell vaccine paradigm described above. Thus, the work presented here covers both computational design and analysis of HIV-derived proteins for potential use as vaccine candidates, and a variety of experimental characterizations of the b12 BNAb and its interactions with HIV gp120 and designed variants thereof.

Chapter 2: Scaffolding the V1/V2 Loop on Heterologous Proteins

As mentioned above, the elucidation of BNAb binding mechanisms is of key importance to recapitulating an antigenic surface of interest, particularly when contemplating the alteration of its context to remove immunodominant competitor epitopes or focus immune responses towards a desired binding mode. To this end, a series of computational designs incorporating the V1/V2 loop region into heterologous proteins were created to enable the structural analysis of the PG9 antibody; these have also been used in a subsequent study to determine the bound structure of the related PG16 antibody[25]. This work, published in *Nature* in 2011[26], was a sizable undertaking for multiple research groups, and included

a wealth of skillful research with which I was not directly involved; Chapter 2 will thus cover my methods and results in greater detail than that afforded by the Nature publication, as well as a discussion of the relevant experimental characterizations performed by the Kwong group and collaborators.

Chapter 3: EPIMAP Modeling of Glycan-Dependent Epitope Accessibility

The RV144 vaccine trial in Thailand is, to date, the most successful human trial at achieving protective efficacy against HIV[27], though its effect was modest and near the threshold of statistical significance. In an effort to glean all available information from the trial's results, investigators at SCHARP conducted a statistical sieve analysis[28] of infecting viral sequences isolated from trial subjects, with a pre-selection of likely sites in the vaccine components to achieve better statistical power for the results. Sergey Menis and I co-designed and implemented a structural modeling software approach as an independent method for picking sieve analysis positions. Our method, labeled EPIMAP, uses atomic-resolution models of antibody binding fragments in conjunction with modeled conformational ensembles of the flexible, heavily glycosylated HIV gp120 vaccine components to rule out sterically inaccessible epitopes on the antigen surface.

Chapter 4: Structural Analyses of Two-Loop b12 Epitope Scaffolds

Efforts to present epitopes of interest outside normally adjacent structural contexts for vaccination typically focus on linear epitopes attached as fusions or insertions into heterologous protein sequences. As many of the most potently targeted HIV *Env* epitopes consist of multiple regions of non-adjacent protein sequence, the development of protein scaffolding methods capable of transplanting complex epitopes onto heterologous proteins while retaining access to their desired, *Env*-like conformations is a key advance in vaccine design. I performed structural characterization of multiple proteins incorporating portions of HIV gp120's functionally conserved CD4 binding site, analyzing the degree of similarity between the designed scaffolds and their gp120 template.

Chapter 5: Biophysical Characterization of b12 And Its Maturation

In an attempt to illuminate the unusual degree of somatic hypermutation commonly observed in anti-HIV BNAbs, we have focused on the often-studied b12 BNAb to analyze at an exhaustive, per-residue level the effects of individual mutations acquired in the process of somatic hypermutation.

Conventional protein interface theory commonly focuses on so-called "hotspot residues" which are necessary for interface formation; to our surprise, no individual point mutation between mature b12 as clinically isolated and its experimentally determined germline precursor sequence is necessary for binding of b12 to gp120. This result is discussed in the context of assumptions prevalent in the HIV field regarding BNAb development, and the ramifications for vaccine design are explored.

Appendix 1: Glycan Masking of Plasmodium vivax Duffy Binding Protein for Probing Protein Binding Function and Vaccine Development

As previously mentioned, much of the current work in the HIV vaccine design field focuses on engineering the surface of *Env* gene products to focus antibody responses towards protective epitopes. N-linked glycosylation is a post-translational protein modification capable of sterically blocking large areas of protein surface with minimal changes to the underlying protein sequence; research performed in collaboration with Joe Smith's research group succeeded in measurably focusing antibody responses against the *Plasmodium vivax* Duffy-binding protein towards protective epitopes[29].

References:

[1] Julien JP et al, *Immunol. Rev.* 2012. 250(1):180-98.
[2] Good MF, *Science* 2013. 341(6152):1352-3.
[3] Yarkoni Y et al, *Immun. Rev.* 2010. 237(1):249-263.
[4] Oracki SA et al, *Immun. Rev.* 2010. 237(1):140-59.
[5] Joller N et al, *Eur J Immunol* 2011. 41(4):889-97.
[6] Haase AT et al, *Nat Rev Immunol* 2005. 5(10):783-92.
[7] Klein F et al, *Science* 2013. 341(6151):1199-1204.
[8] Wei X et al, *Nature* 2003. 422(6929):307-12.
[9] Klein F et al, *Science* 2013. 341(6151):1199-204.
[10] Finton KAK et al, *PLoS Pathogens* 2013. 9(9):e1003639.
[11] Haynes BF et al, *NEJM* 2012. 366:1275-86.
[12] Xiao P et al, *J Virol* 2012. 86(8):4644-57.
[13] Roben P et al, *J Virol* 1994. 68(8): 4821-4828.
[14] Barbas CF et al, *PNAS* 1992. 89:9339-9343.
[15] Barbas CF et al, *PNAS* 1992. 89:9339-9343
[16] Muster T et al, *J Virol* 1993. 67(11):6642-6647.
[17] Zwick MB et al, *J Virol* 2001. 75(22):10892-10905.
[18] Keller PM et al, *Virology* 1993. 193(2): 709-716.
[19] Trkola A et al, *J Virol* 1996. 70(2):1100-1108.
[20] Wu X et al, *Science* 2010. 329(5993):856-861.
[21] Walker LM et al, *Science* 2009. 326(5950):285-289.
[22] Wu X et al, *Science* 2010. 329:856-61.
[23] Walker LM et al, *Science* 2009. 326(5950):285-289.

[24] Huang J et al, *Nature* 2012. 491(7424):406-12.
[25] Pancera M et al, *Nat. Struct. Mol. Biol.* 2013. 20(7):804-813.
[26] McLellan JS et al, *Nature* 2011. 480(7377):336-43.
[27] Rerks-Ngarm S et al, *NEJM* 2009. 361(23):2209-2220.
[28] Rolland et al, *Nature* 2012. 490(7420):417-420.
[29] Sampath S et al, *PLoS Pathogens* 2013. 9(6):e1003420.

Chapter 2: Scaffolding the gp120 V1/V2 Loop on Heterologous Proteins

Introduction

The recent discovery of the PG9 and related PG16 broadly neutralizing antibodies (BNAbs) against conformation-dependent epitopes on HIV gp120[30] heralded a new phase in the quest to elicit protective BNAb responses against HIV; these antibodies succeeded for the first time in neutralizing the majority of identified HIV primary isolates at inhibitory concentrations reasonably elicited by vaccination. Previously isolated BNAbs either inhibit ~50% or fewer primary isolates (b12) or do so at concentrations too high to be feasibly induced by a vaccine. Additionally, the regions of the variable loops targeted by PG9 and PG16 are known to be common targets of clinically observed anti-HIV antibody responses, in stark contrast to the gp120 CD4 binding site and gp41 MPER targets of b12 and 4e10 respectively.

The state of knowledge regarding PG9's epitope on gp120, prior to the beginning of the present work, was that it involved residues located in the flexible V1/V2 loop region; this region is located on the far end of the *Env* trimer surface from the viral membrane, and is thought to associate between trimer subunits. Early failures to identify gp120 monomers capable of binding PG9 led to the conclusion that its epitope must accordingly cover adjacent trimer subunits.

Development of designed immunogens to more reliably elicit neutralizing responses against V1/V2 loop targets has historically been hindered by the lack of available structural information regarding gp120's variable loops. Extant crystal structures of gp120 have truncated the V1/V2 loop[31] or failed to place it within well-defined density; similarly, structures of the *Env* trimer determined by cryoelectron microscopy have failed to place a sufficient subset of the loop for computational placement of the remainder, though some stable density corresponding to a "cap" is visible.

Epitope scaffolding is a protein design methodology for transplanting antibody epitopes of interest onto heterologous proteins; initial efforts in epitope scaffolding focused on structurally defined linear epitopes, particularly on *Env*, with the aim of presenting them in an immunologically novel context to elicit protective anti-*Env* antibody responses. While originally developed to remove immunodominant or distracting epitopes for the eventual use of epitope-bearing "scaffold" proteins in vaccines, scaffolds often have superior biophysical properties for structural comparison as compared to the original antigenic protein bearing the unscaffolded epitope in its native context. Thus, we designed a panel of scaffold

proteins presenting the V1/V2 loop in as close to a native conformation as possible given limited structural information with the goal of allowing native-like binding of the V1/V2 loop by PG9 in a crystallizable complex.

Results

A V1/V2 Deletion Stub Provides a Structural Surrogate for Scaffold Design

The computational design methods used in past epitope scaffolding efforts[32] require an atomic-level model of complete backbone structure for proper selection and modification of potential scaffold proteins. Given the relatively unstructured nature of our V1/V2 target sequence, a short sequence from a V1/V2-truncated gp120 construct used for previous crystal structure determination was selected as an insertion surrogate; this ensured that the termini of the V1/V2 insertion would be as close as possible to their native conformation in gp120.

A disulfide bond linking the ends of a CVGAGSC motif in the 1RZJ crystal structure forms a usefully well-defined structural element suitable for surrogate insertion, as the disulfide linkage is tightly geometrically constrained and provides some assurance that the local protein geometry will be similar for the full-length V1/V2 and the scaffold proteins. For selection of scaffolds, the cysteines from the 1RZJ stub and two flanking residues per terminus were used to generate matches (Fig. 2.1); these matches were subsequently filtered by a 0.5-Å root-mean-squared deviation (RMSD) cutoff.

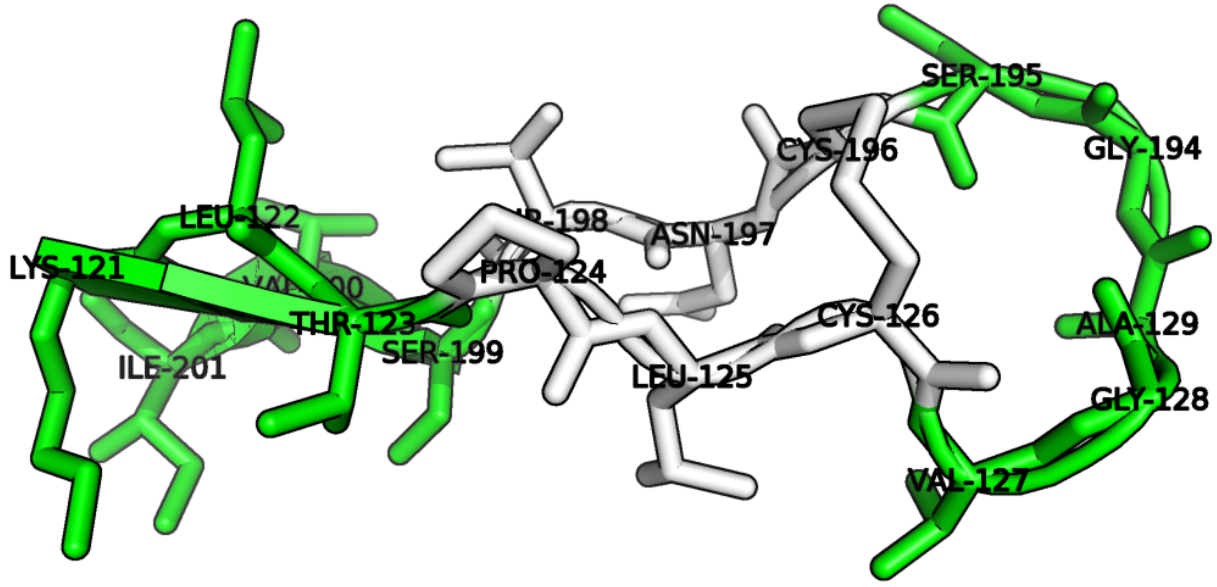


Figure 2.1: The 1RZJ V1/V2 Loop Stub Structure. Residues 124-126 and 196-198 were used to generate computationally aligned structural matches to heterologous proteins, allowing the truncated V1/V2 loop stub (residues 127-195; HXB2 numbering) to remain unaltered in the modeled insertions.

Scaffolds were selected in two batches; first, a selection of short (<100 amino acids) proteins used for prior NMR structural determination to offer an alternate experimental method; and second, a selection of larger (135-741 amino acids) proteins expressible in human cell culture and previously crystallized at high resolution, intended to offer a reliably stable platform for integration of the putatively structured V1/V2 loop while leaving sufficient well-ordered intermolecular contacts for successful crystallization. The curated outputs of these batches were advanced to computational design.

Fixed-Backbone Design Ameliorates Introduced Structural Pathologies

Matching results from the initial stage were subjected to fixed-backbone protein design in Rosetta; flexible-backbone design, as previously used to remove flaws in scaffold geometry introduced by epitope placement, was unnecessary due to the high-quality matches available given the common beta-hairpin geometry used in the 1RZJ V1/V2 stub. Points of particular concern were the removal of steric clashes between adjacent protein side chains and the V1/V2 base disulfide, and the amelioration of voids formed in the hydrophobic core packing of designs based on the extension of the protein surface by the V1/V2 insertion. Additionally, scaffold proteins of known function were mutated to inactivate preexisting binding and catalytic sites; this was particularly important for the 1FD6 scaffold, which has a native immunoglobulin-binding function inimical to passive presentation of a loop for antibody co-crystallization. 17 total designs passed all computational stages; 11 derived from small, NMR-determined proteins, and 6 from larger, previously secreted proteins (Fig. 2.2, Table 2.1).

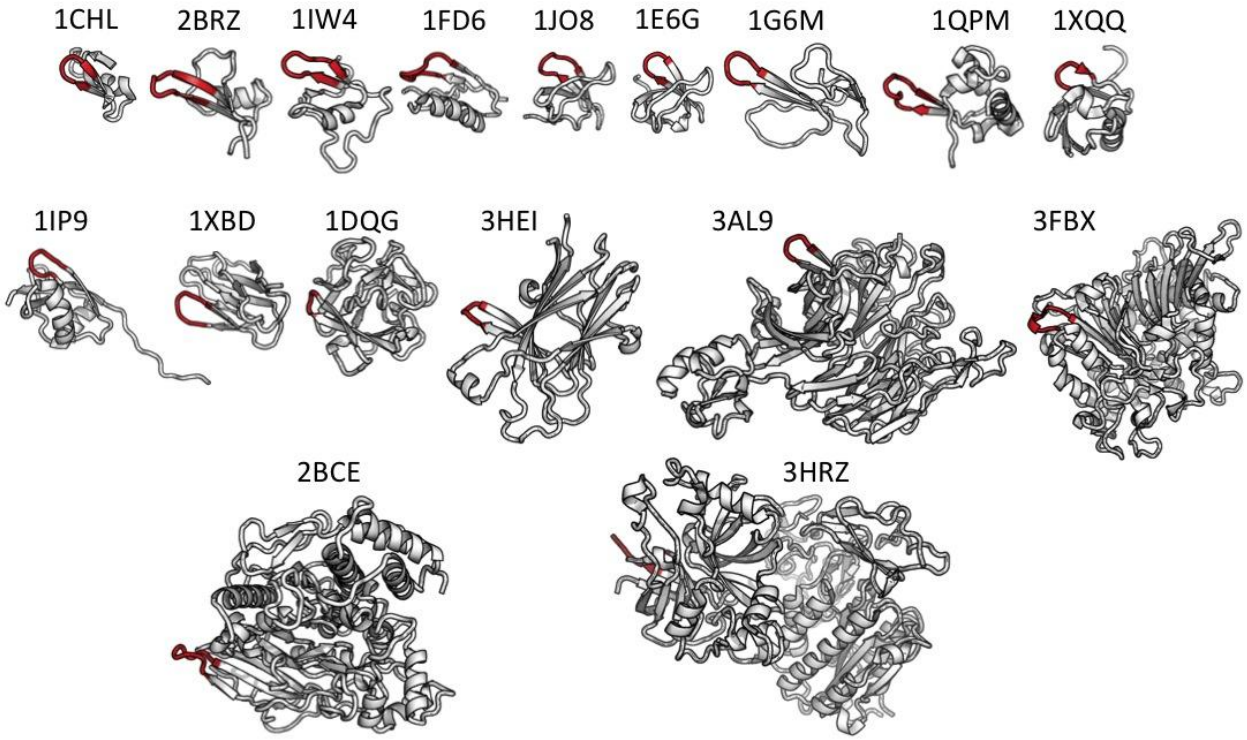


Figure 2.2: 17 Scaffolds Selected for Experimental V1/V2 Insertion Testing. Scaffolds are displayed with regions excised for V1/V2 sequence insertion highlighted in red; orientations are derived from the shared 1RZJ-derived stub alignment, and scale is constant.

V1/V2 Insertion for Scaffold Screening

Finished computational designs had their stub surrogate sequences replaced with a V1/V2 loop sequence from the gp120 encoded by the YU2 strain of HIV, which had previously been identified by tight binding to PG9. Of the six scaffolds from the larger parent proteins, only the shortest expressed in HEK293F cells, and exhibited poor solution behavior. The smaller protein scaffold set yielded three constructs with measurable binding to a panel of murine monoclonal anti-gp120 antibodies; two (1JO8 and 1FD6) also demonstrated binding to α4β7 integrin, known to bind gp120 via the V1/V2 loop. Integrin binding by the V1/V2 loop constitutes a known facet of the *in vivo* function of gp120, and accordingly served as a measure of the physiological relevance of the scaffold-generated V1/V2 conformational ensemble.

Table 2.1: V1/V2 Scaffold Expression

Source PDB ID	Length, Pre-Insertion	V1/V2 Insertion Residues	HEK293 Expression	Anti-V1/V2 Ig Binding	α4β7 Integrin Binding
1CHL	36	30-31	Yes	0/5	
2BRZ	54	39-45	Yes	0/6	
1IW4	55	25-30	Yes	1/6	
1FD6	57	9-15	Yes	6/6	Yes
1JO8	58	42-43	Yes	6/6	Yes
1E6G	62	46-49	Yes	6/6	No
1G6M	62	32-33	Yes	1/6	
1QPM	69	38-44	Yes	4/6	
1XQQ	76	46-47	Yes	5/6	
1IP9	85	20-21	Yes	0/6	
1XBD	87	53-59	Yes	0/6	
1DQG	135	19-23	Yes	Not Tested	
3HEI	160	30-37	No		
3AL9	539	255-258	No		
3FBX	559	183-188	No		
2BCE	579	19-26	No		
3HRZ	741	444-449	No		

Non-YU2 gp120 sequences were screened for binding to PG9 by monomeric gp120, as opposed to conformationally induced binding in the context of the *Env* trimer; sequences TRJO from HIV clade B, as well as CAP45, ZM53 and ZM109 from clade C, yielded high-affinity gp120:PG9 binding. Affinities for the ZM109 insertion sequence as presented by the 1FD6 scaffold were within an order of magnitude of those characterizing the ZM109 gp120:PG9 interaction.

Purification of sufficiently homogeneous complex for crystallization required an on-column binding protocol wherein heterogeneously glycosylated scaffold proteins produced in HEK 293 cells were bound to PG9 IgG; Fab:scaffold complexes were generated by proteolytic cleavage of column-immobilized IgG and subsequent size-exclusion chromatography. Combinatorial crystallization attempts for both the 1FD6

and 1JO8 scaffolds with loop sequences from TRJO, 16055, CAP45, ZM53 and ZM109; crystals were successfully grown from the 1FD6:CAP45 scaffold:loop construct and a variant of the 1FD6:ZM109 construct with four N-linked glycan sites mutated to alanines.

Crystal Structures of PG9:Scaffold Complexes Reveal Glycan-Mediated Binding

The 1FD6:CAP45 and 1FD6:ZM109 crystal structures jointly reveal a highly structured motif in the V1/V2 loop as presented by the scaffold. A relatively unusual structural feature characterizing both complexes is the presence of a large, crystallographically stable N-linked glycosylation occupying and apparently partially mediating the antibody:loop interface. While the protein:protein component of the interface is exclusively mediated by the protruding heavy chain CDR3 of PG9 (Fig. 2.3A, 2.3C), the light chain and other heavy chain CDRs form a more canonically antibody-like concave binding surface to interact with a V1/V2 glycan covalently linked to gp120 position 160 (Fig. 2.3B).

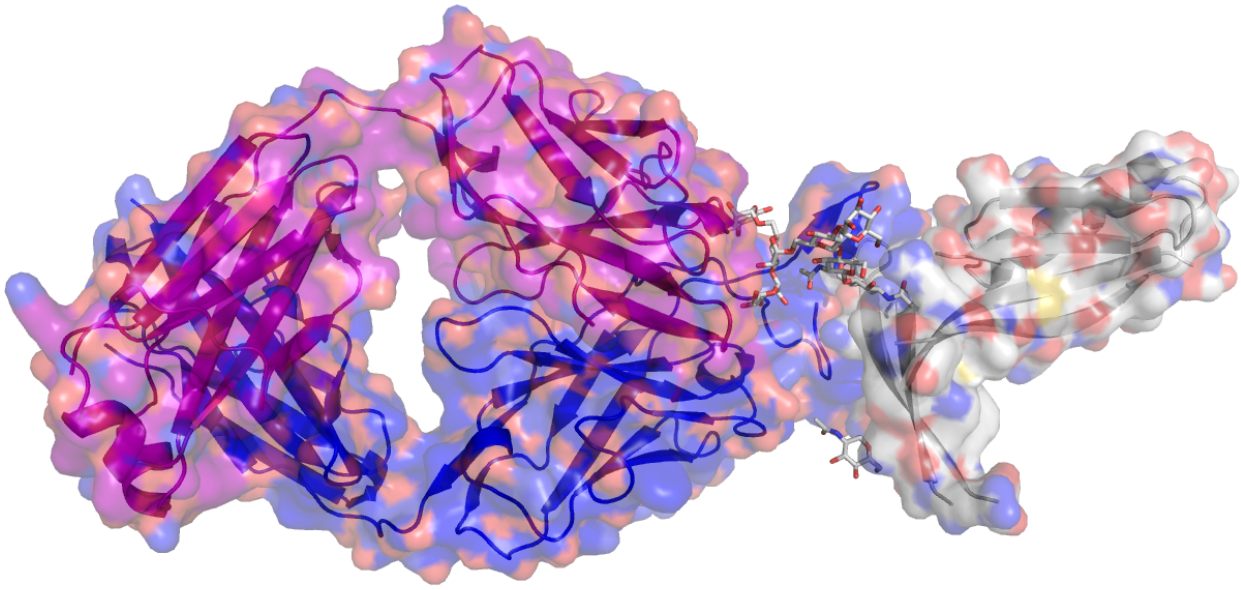


Figure 2.3A: Overall Structure of the 3U2S 1FD6_ZM109:PG9 Scaffold:Fab Complex. The PG9 heavy (blue) and light (purple) chain bind to a 1FD6-derived scaffold incorporating and displaying the V1/V2 loop sequence from the ZM109 strain of HIV gp120.

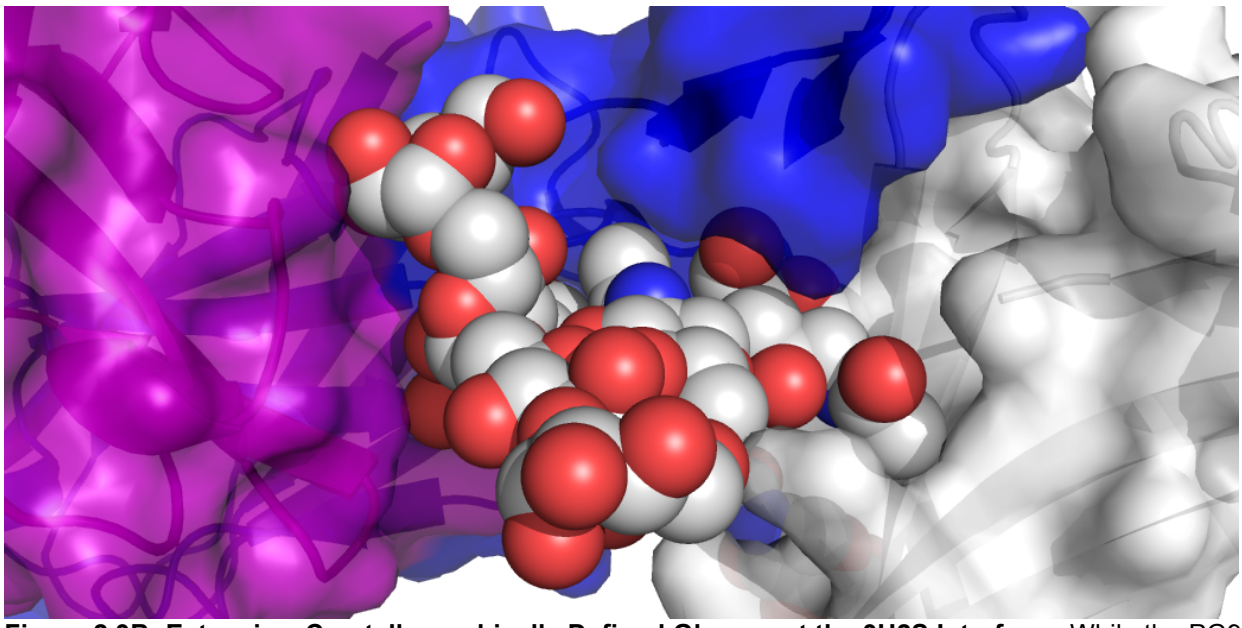


Figure 2.3B: Extensive, Crystallographically Defined Glycans at the 3U2S Interface. While the PG9 heavy chain solely mediates protein:protein contacts, the light chain CDRs contact the observed N-linked glycan structure shown here, which is covalently linked to the asparagine residue at V1/V2 position 160.

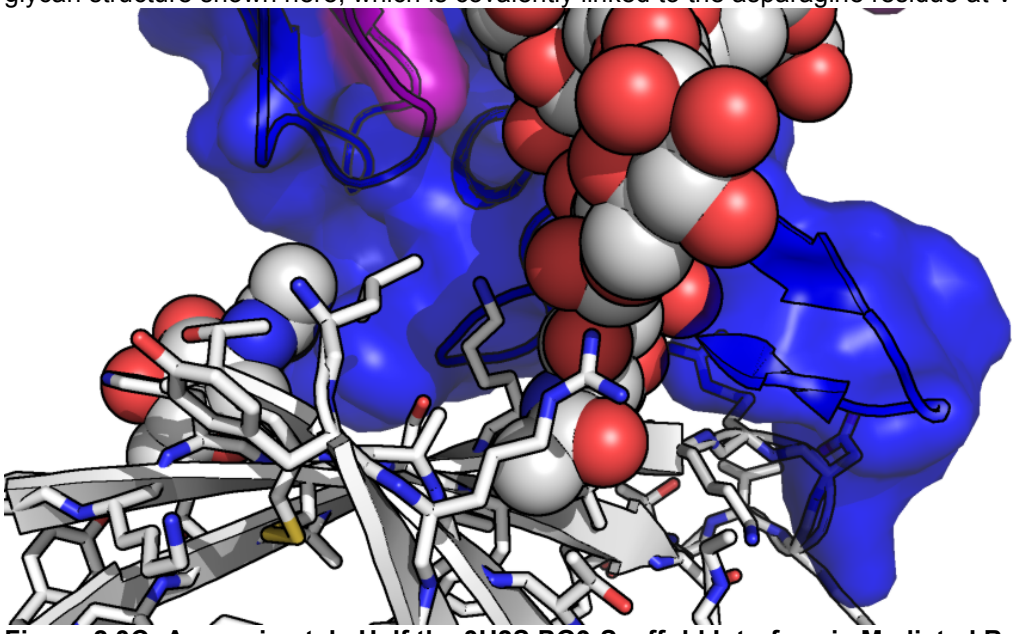


Figure 2.3C: Approximately Half the 3U2S PG9:Scaffold Interface is Mediated By Glycans. Non-glycan contacts are restricted to the region of protein sequence between the N-linked glycan sites at V1/V2 sequence positions 160 and 173.

Protein:protein contacts at the PG9:V1/V2 interface are primarily between the PG9 CDRH3 loop and residues 167-172 in the V1/V2 loop (Fig. 2.4A). The PG9 CDRH3 forms a network of hydrogen bonds between its backbone atoms and these residues; additionally, tyrosine sulfation sites at residues 100G and 100H are most likely interacting with the positively charged side chains at V1/V2 positions 169

(lysine) and 168 (arginine) (Fig. 2.4B). While neither V1/V2 position is strictly conserved, residue 168 is typically a positively charged lysine or arginine.

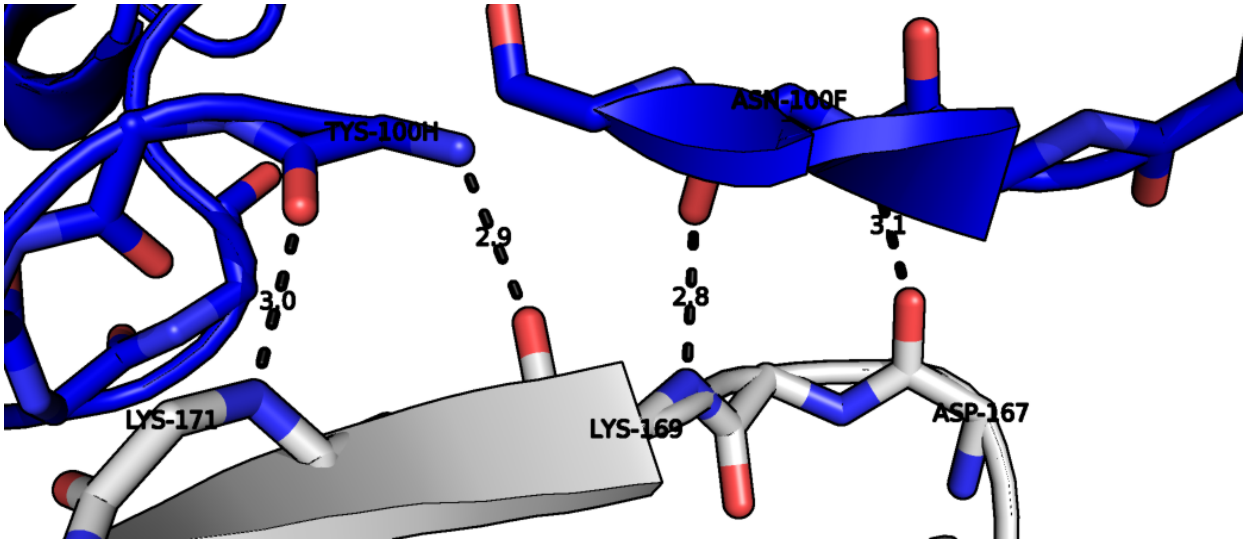


Figure 2.4A: Backbone:backbone Hydrogen Bonding at the PG9:V1/V2 Interface.

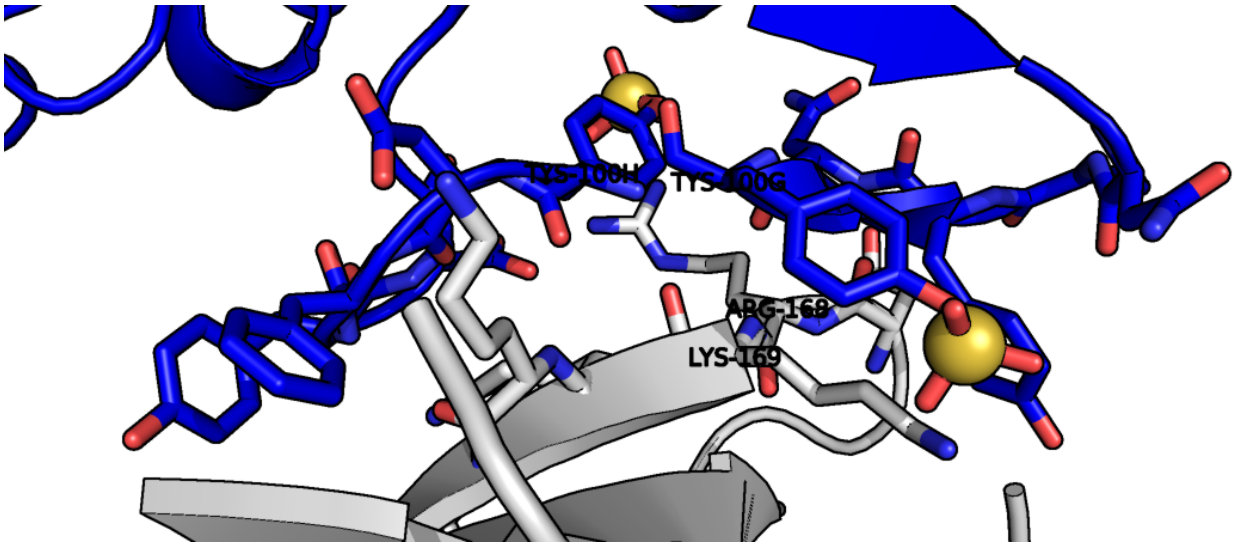


Figure 2.4B: Tyrosine Sulfation and Positively-Charged V1/V2 Residues. Side chains of resides K169 and R168 interact with PG9 sulfotyrosine residues, and are commonly present in the sequences of primary clinical isolates of HIV.

Comparison of V1/V2 Scaffold Structures with Design Models

With the experimentally determined structure of the V1/V2 loop as seen in the 1FD6:PG9 complexes, evaluation of the original design models becomes possible. Alignment of the V1/V2 loop structure with the design models for the scaffolds (Fig. 2.5) reveals that all expressible scaffolds feature ample room for the V1/V2 loop to attain its PG9-bound conformation in solution without steric interference

from the scaffold. The designed proteins with the best binding activities (1FD6, 1JO8, and 1E6G) also appear to offer additional contacts buttressing the V1/V2 insertion (as experimentally observed in the 1FD6-based structures). The 1IP9 model also suggests significant additional scaffold:loop contacts, suggesting that the requirement for non-interference with binding supersedes the energetic benefits of extra contacts.

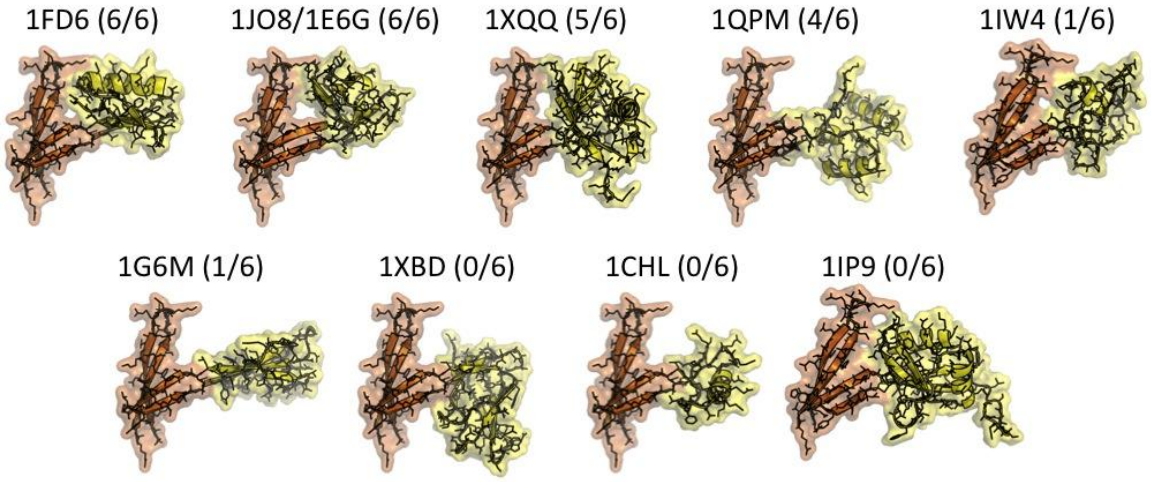


Figure 2.5: Experimentally Determined V1/V2 Structure and Original Design Models. All ten tested scaffolds (yellow) allow the V1/V2 loop (orange) to achieve its PG9-bound conformation; the 1FD6, 1JO8, 1E6G, and 1IP9 scaffolds allow additional supporting contacts.

Discussion

From a protein design perspective, scaffolding the V1/V2 loop posed an unusual challenge. Canonical protein design methods emphasize a target structure, or ensemble of related structures, which the designed proteins are to achieve; computational structure modeling is used to enable closer adherence to the template model. With the V1/V2 loop as a structural target, the design method is forced to leverage minimal extant information to acquire an improved model via experiment, which could in turn be used as the basis for more typical epitope scaffolding. Successful experimental validation of the designs, then, represents an expansion of the problem set for the methods rather than reaffirming their validity for more canonical uses.

An important aspect of the unusual nature of this design project lies in the evaluation of the resulting constructs' structures for potential improvement of the underlying design methods. The observed crystal structures of the 1FD6 scaffold variants exhibit only minor differences relative to the

computational model within the retained sequence outside the V1/V2 insertion; the junction between the V1/V2 loop and the flanking scaffold residues exhibits no major structural pathologies (data not shown). The more interesting insights come from the superior binding of scaffolds which offer secondary contacts with the stable V1/V2 insertion, suggesting not only that scaffolds themselves enhance PG9 binding by favoring the bound loop conformation, but that gp120 monomers - and, potentially, trimers - may also vary in affinity for PG9 via a similar mechanism.

Structural insights into PG9 binding achieved through the use of the scaffolds described here center around the antibody's use of conserved glycan residues and main-chain atom hydrogen bonds to circumvent many possible escape mutations; side chains in most of the V1/V2 loop are simply in insufficient contact with the antibody CDRH3 to sterically hinder binding sufficiently to disrupt the protein:glycoprotein interface. Thus, the crystal structure offers a mechanistic explanation for the extreme breadth of binding demonstrated by PG9; despite a target epitope located in a relatively flexible and mutationally variable region of protein structure, the bound state targeted by the antibody is essentially invariable across the sequence space of infectious HIV sequences.

Materials and Methods

Protein Scaffolding Software

The Multigraft Match software mode in the Rosetta protein structure prediction package uses iterative placement of a linear region of protein structure model (the "graft") within a heterologous acceptor protein (the "scaffold"), with subsequent scoring of the proposed result model by the Van der Waals terms of Rosetta's scoring function to identify graft placements with significant steric clash. Previously published work using Epigraft has followed the matching phase with a "grafting" phase, in which the protein backbone in the connecting regions of sequence between graft and scaffold is varied to minimize torsional, steric and hydrogen-bonding pathologies in the model. In the V1/V2 loop insertion case, inefficiencies in Rosetta's flexible-backbone sampling particular to the extended secondary structure characteristic of paired antiparallel beta strands such as the V1/V2 base led to the use of low-RMSD matches with backbone flexibility used only to check that closure was possible rather than using the closed model for subsequent fixed-backbone design procedures.

Potential Scaffold Data Sets

Multigraft Match uses, as input, curated protein structure files in the Protein Databank (PDB) format. PDB files are pre-treated by the removal of non-protein atoms, non-canonical residues, secondary conformations, and non-atom records such as experimental conditions for structural determination. The resulting all-amino-acid PDB database (PDBAA) is organized by two-letter code to minimize potential directory access slowdowns in the storing file system. Typically, crystal structures of proteins of at most 250 amino acids per chain, monomeric solution state, and resolution equal to or better than 2.5Å are used for scaffold design due to generally favorable handling properties.

For scaffolding the V1/V2 loop, the PDBAA generation process was applied to two subsets of experimentally determined structures. For the 11-member small protein cohort of designs, NMR-determined structures - typically corresponding to multiple "MODEL" records denoting a conformational ensemble or solution range for NMR constraints rather than the single model with variable-multiplicity atoms characteristic of crystal structures - were processed into individual files on a per-model for subsequent PDBAA_NMR records. The larger scaffold set was processed by querying the PDB for expression organism (homo sapiens) and removing the typical protein length cutoff; oligomeric state and resolution thresholds were retained to ensure crystallizability.

Experimental Characterization and Crystallography

As I did not participate in the experimental validation of these designs, nor see descriptions of the methods used outside of the publication process for the eventual paper manuscript, please refer to that paper for details regarding experimental procedures.

Acknowledgements

The experimental work conducted using these V1/V2 scaffold designs was performed by researchers in and collaborating with Peter Kwong's research group at the National Institutes of Health; particular credit goes to Jason McLellan and Marie Pancera for the purification and crystallization of the scaffold:PG9 complexes, and Mallika Sastry for the initial expression and binding assay work to identify candidates for subsequent follow-up.

References:

[30] Walker LM et al, *Science* 2009. 326(5950):285-9.
[31] Zhou T et al, Nature 2007. 445(7129):732-7.
[32] Correia BE et al, *Structure* 2010. 18(9):1116-26.

Chapter 3: Modeling Glycan-Dependent Epitope Accessibility

Introduction

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

B cell epitope usage (BCEU), often simplified in the literature as antibody epitope usage, may be defined as the targeting of surfaces present on an antigen of interest, particularly in a differential manner leading to increased immune responses against a subset of the antigen's structure. Such differences in BCEU are most easily observed via so-called immunodominant epitopes which are preferentially targeted to an extent which prevents efficient targeting of other potential epitopes, often providing a selective advantage for the immunodominant antigen's parent organism by directing B cell responses away from functionally required sites of conserved sequence towards hypervariable antigen subsequences more amenable to immune escape.

Reliable BCEU prediction for antigens of interest would, in theory, allow efficient targeting of more immunologically useful epitopes by the attenuation of immunodominance. However, the vast structural heterogeneity of BCRs which forms the basis for antibody recognition and the intrinsically random selection process by which B cell clones compete for proliferative signaling have, to date, rendered BCEU prediction per se intractable at a level useful for vaccine development. Current best-practice protocols in BCEU prediction typically use solvent exposure of antigen residues (CEP[33]), often adding statistical information derived from non-structural experiments (DiscoTope[34], BePRO/PEPITO[35]); antibody shape complementarity using representative antibody probes offers additional information (Ellipro[36]). Even the best BCEU prediction methods exhibit positive identification accuracies below 30%, however[37], likely reflecting the underlying biochemical feasibility of multiple sites on antigen surfaces as epitopes and the confounding issues with B cell selection.

If we lack ability to discriminate between BCEU propensity for epitopes on an antigen of given structure, then, a potential workaround for conformationally variable antigens is to exploit conformational flexibility to rule out certain epitopes as less accessible or inaccessible to antibodies. This is of particular value in determining epitope usage for viral glycoproteins, which bear self-like sugars unlikely to generate strong B cell responses except in conjunction with adjacent protein surface; glycoproteins are also known

to be highly conformationally variable due to the intrinsic flexibility of their covalently attached glycan groups. The goal of the work contained herein is thus to computationally describe the limitations placed on BCEU for the HIV-1 gp120 protein due to its high conformational variability and glycosylation.

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

A common feature of the eukaryotic secretome is the post- or cotranslational modification of proteins by the covalent addition oligosaccharide groups. N-linked glycosylation chemically modifies surface asparagine residues found in an NX(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[38]; the process of enzymatic glycan modification results in a heterogeneous population of glycoprotein species sharing an identical protein sequence.

While antibodies against non-human glycosylation structures are observed to play an important protective role against parasitic eukaryotic pathogens, viral proteins display human-type glycosylation as a consequence of their synthesis using hijacked 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.

HIV-1 gp120 is notable for its high density of exposed N-linked glycosylation sites (Fig. 3.1). Current progress in structural determination of the HIV-1 Env trimer - largely composed of gp120 - supports the hypothesis that large areas of the trimer surface are spatially occluded by glycosylation. Indeed, protein sequence mutations affecting glycosylation have repeatedly been identified as important for binding and neutralization sensitivity by antibodies directed against gp120, and broadly neutralizing antibodies against gp120 tend to target deglycosylated patches of protein surface. Additionally, certain areas of gp120 express self-like glycans in sufficient density to present non-self-like glycan epitopes - either by simple proximity to an extent unseen in the self glycoprotein repertoire used to enforce B cell tolerance, or by the generation of non-self-like glycans via steric inhibition of glycan-modifying enzymes by the abnormally high local density of sugar residues.

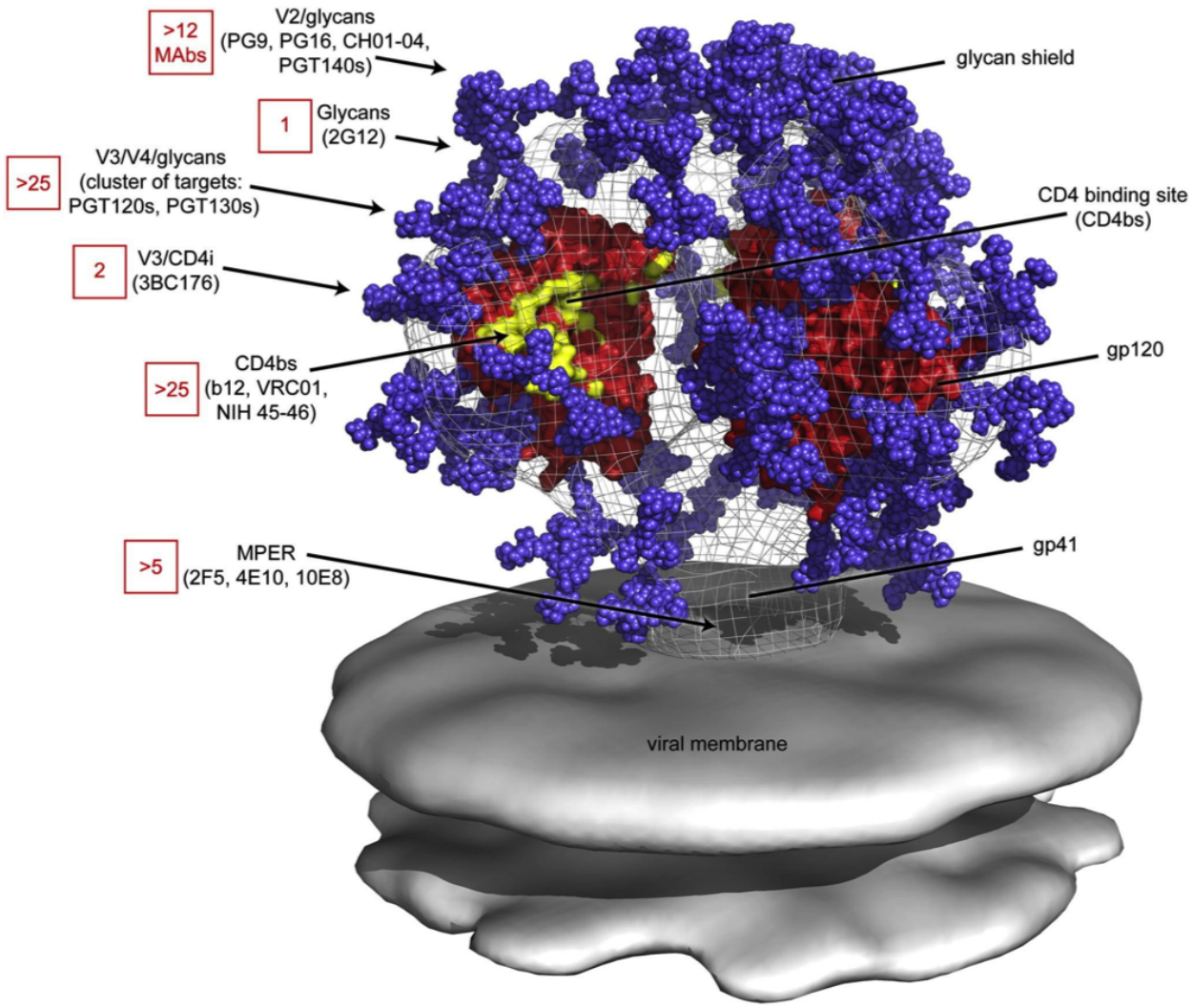


Figure 3.1: Glycosylated gp120 Models Docked into a Cryo-Electron Microscopy Structure of the HIV *Env* Trimer. Epitopes used by broadly neutralizing antibodies are noted; glycans corresponding to missing V1/V2 loop density are placed above trimer cap. Figure is reproduced here courtesy of Dennis Burton[39], and is an annotated update of a figure originally created by Bill Schief[40]; permissions have been obtained from Elsevier and Dennis Burton.

Vaccine Sieve Analyses

Establishing the comparative degree of accessibility of potentially blocked epitopes is a major concern in vaccine design, particularly in the case of heavily glycosylated viral envelope proteins such as the HIV-1 Env and influenza hemagglutinin trimers. While we expect the method discussed herein to have considerable utility in evaluating potential vaccine components, the immediate requirement which precipitated the method's development was the need to limit a sieve analysis of the RV144 vaccine trial to residues expected to be antibody-accessible.

The RV144 trial represents an important milestone for the HIV vaccine field in that, for the first time, modest protection was observable 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-on research has attempted to disentangle potential mechanisms by which RV144 achieves protection; a particularly useful tool for post-hoc analysis of the vaccination response is the identification of statistical correlates of protection[41] among parameterizable 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[42]; IgA response against gp120 was a negative correlate of protection.

Immune correlates of protection are not, of 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 increases with the number of positions considered. As such, methods for 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.

Results

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 of sites on a target glycoprotein's surface. This rests on the assumption that some combination of CDRs will be capable of forming an energetically favorable, high-affinity interface with any particular "patch" of glycoprotein surface residues and sufficiently close antibody placement, provided steric hindrances by glycosylation and flexible regions of the target can be overcome.

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. 3.2, left-hand panels); 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. 3.2, center panels), 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. 3.2, right-hand panels)

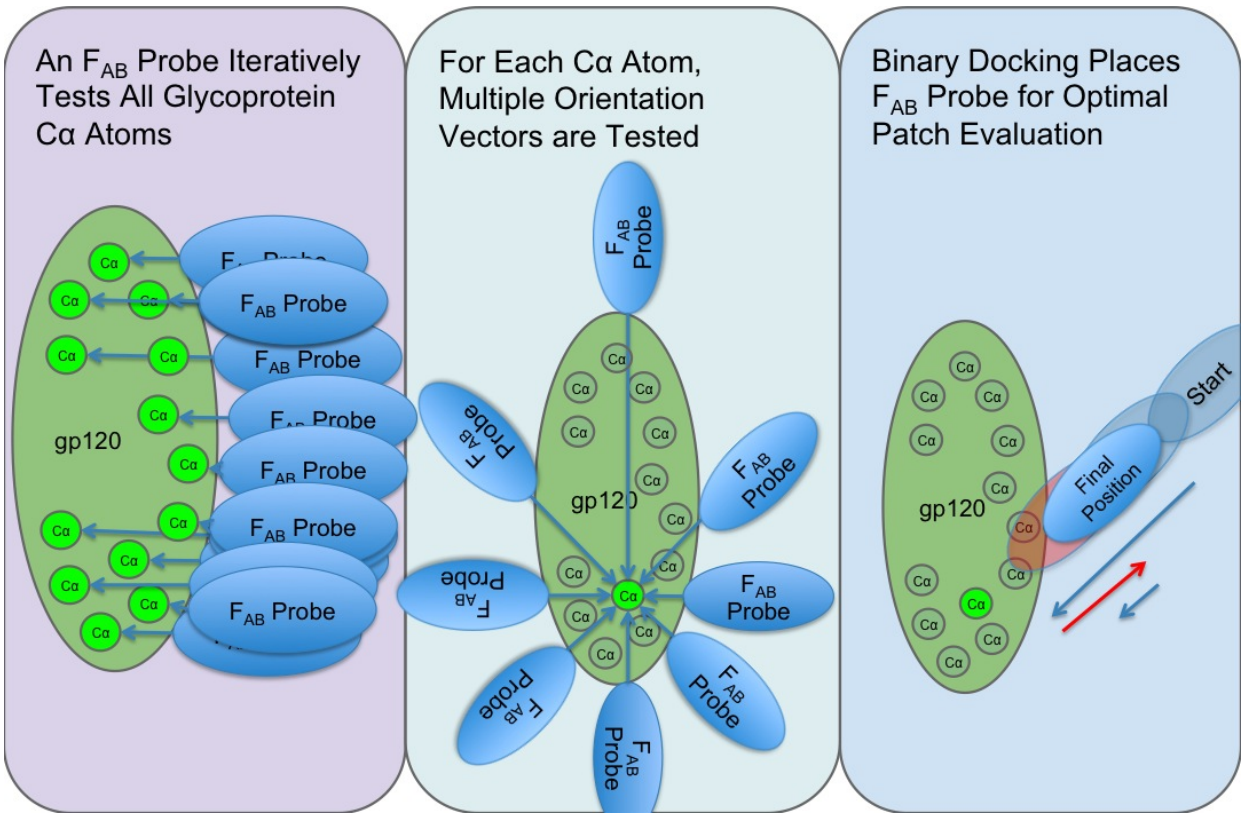


Figure 3.2A: Schematic Representation of Antibody Placement Procedures. A F_{AB} probe model is placed in multiple orientations (center panel) surrounding each Cα of the target glycoprotein model (left panel); optimal positioning of each probe orientation is achieved by a binary-search docking procedure (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).

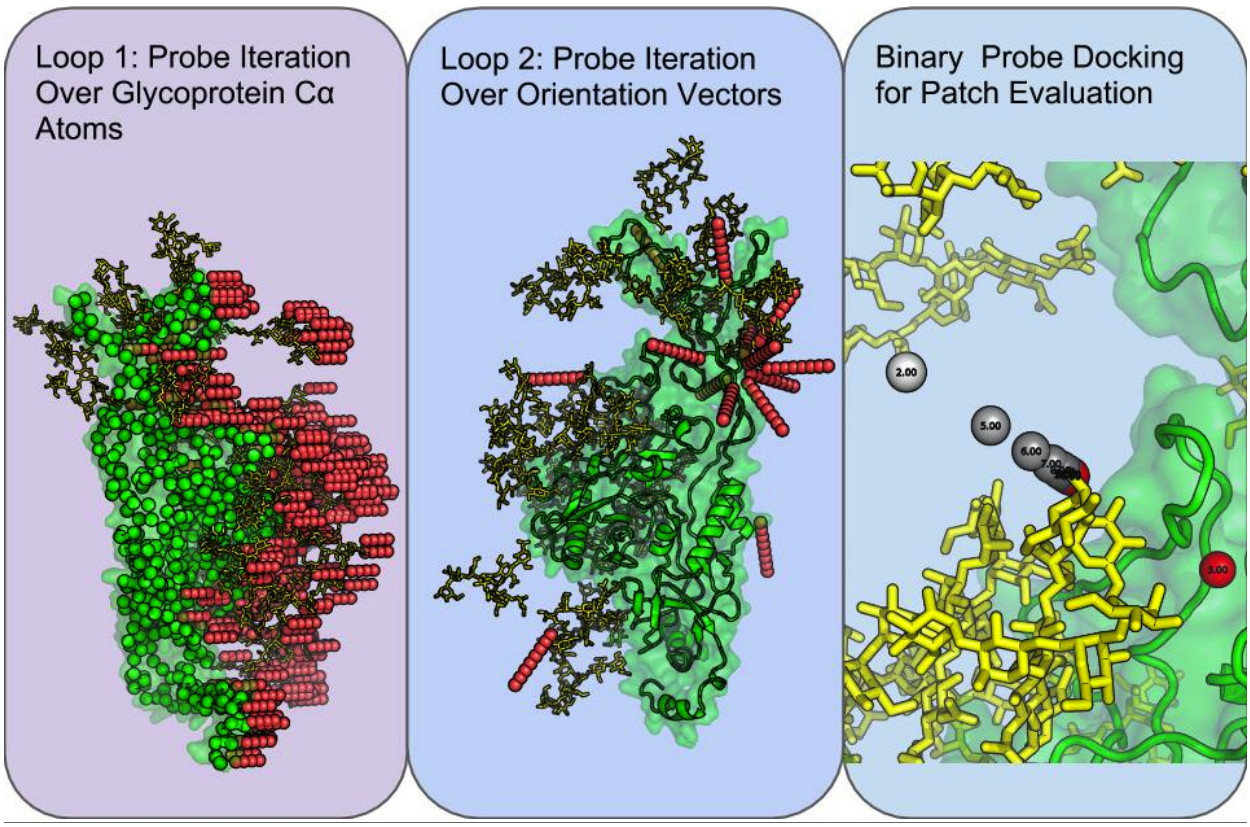


Figure 3.2B: Glycoprotein Representation of Antibody Placement Procedures. For these figures, the red ball series represents a larger F_{AB} probe. Note that glycans (yellow) may sterically block some probes from accessing the protein surface (green). In the right-hand panel, steps in the binary search procedure are ordinally numbered; following step 2 (off-white), step 3 (red) is placed within the target glycoprotein model, resulting in the probe "backing off" of the glycoprotein surface and re-approaching through iteratively smaller movements.

Evaluation of antibody placements consists of an initial steric clash check to identify physically impossible interaction models, and subsequent identification of potentially interacting residues of the target glycoprotein as scored by proximity to a subset of the query Fab model corresponding to the CDRs. A key refinement in our code is the implementation of a spatially indexed binning system for the viral glycoprotein model used in a particular program instance; this allows steric checks, which are at heart low-distance cutoff conditions, to be applied only to nearby atoms in the query Fab model. As the Fab model also 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 re-initializing or updating glycoprotein atom positions.

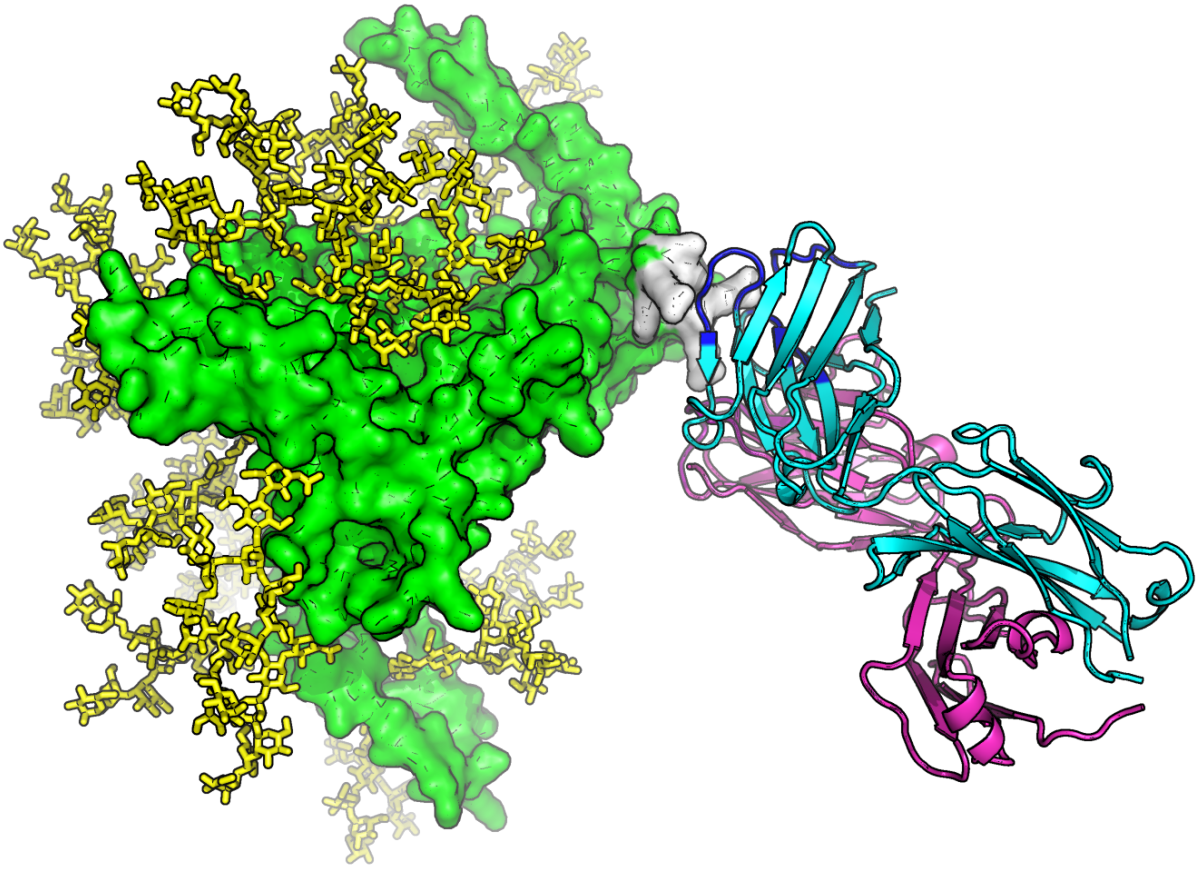


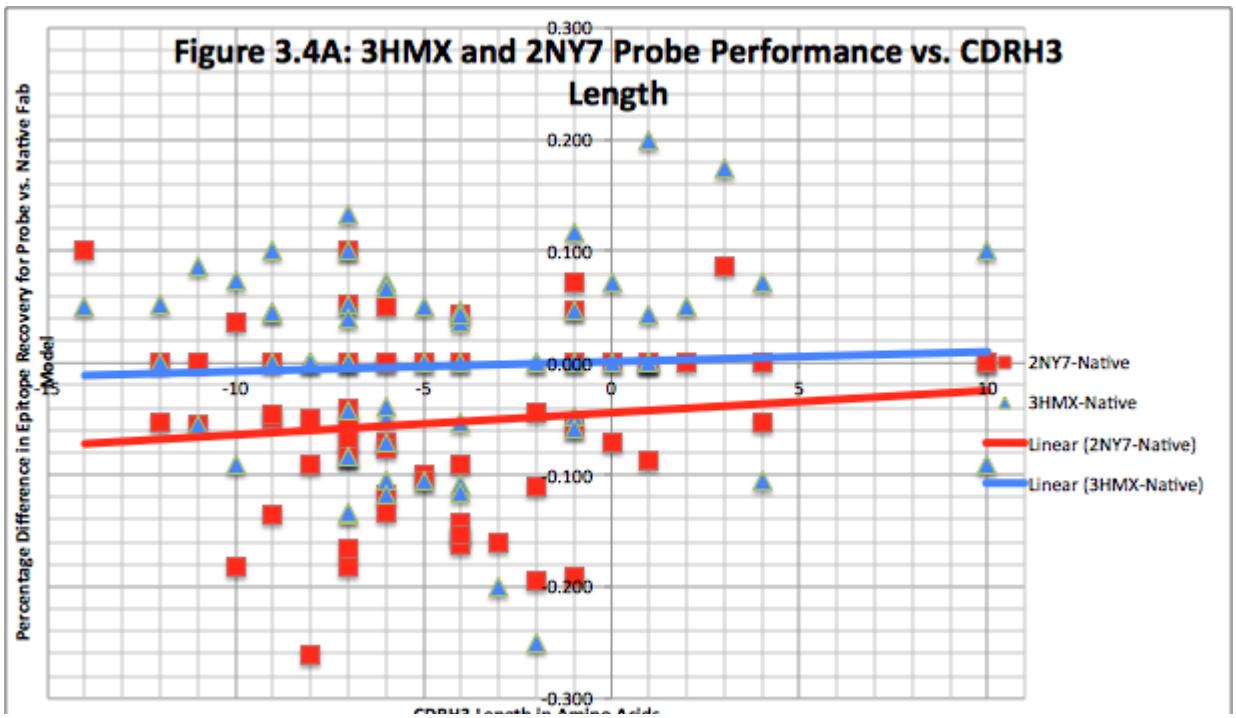
Figure 3.3: Antibody Patch Definition. The 3HMX antibody probe (cyan heavy chain, magenta light chain) is positioned adjacently to the gp120 (green) surface; a "patch," (white) defined as all glycoprotein atoms within 6Å of a designated interaction subset of the antibody corresponding to the antibody CDRs (blue heavy chain, purple light chain). No glycans (yellow) participate in this patch.

Fab Models Are Interchangeable for Patch Accessibility Prediction

Initial patch prediction attempts were performed on a glycosylated gp120 model using a structural model of the anti-HIV neutralizing antibody b12[43] Fab construct; this provided recapitulation of the known b12 epitope as a method viability metric intrinsic to the data set. As a subsequent check of the b12 model's suitability for patch prediction, a benchmark protocol was performed over a set of high-resolution Fab:antigen complex crystal structures (Table 3.1, located as an appendix to this chapter).

Comparison of epitope recapitulation by b12 or the native Fab model for each target complex revealed generally similar performance between our surrogate and native models, further supporting our initial results[44] obtained with the b12 model. Investigation of the cases where b12's performance dropped significantly (15% or greater loss of recapitulation) relative to the native Fab for a particular target

revealed a trend towards shorter CDRH3s for poor-performance cases (fig. 3.4A). A possible mechanism for this result is the steric inability of b12 to place sufficient CDR residues within the proximity cutoff of a target epitope required for proper scoring, whereas short-CDR Fab models will more commonly be placed in a correct orientation but may generate false negatives for residues buried in loop-accessible cavities within a larger epitope. This appears to be the case in cases such as 3UY7 and 3HI1.



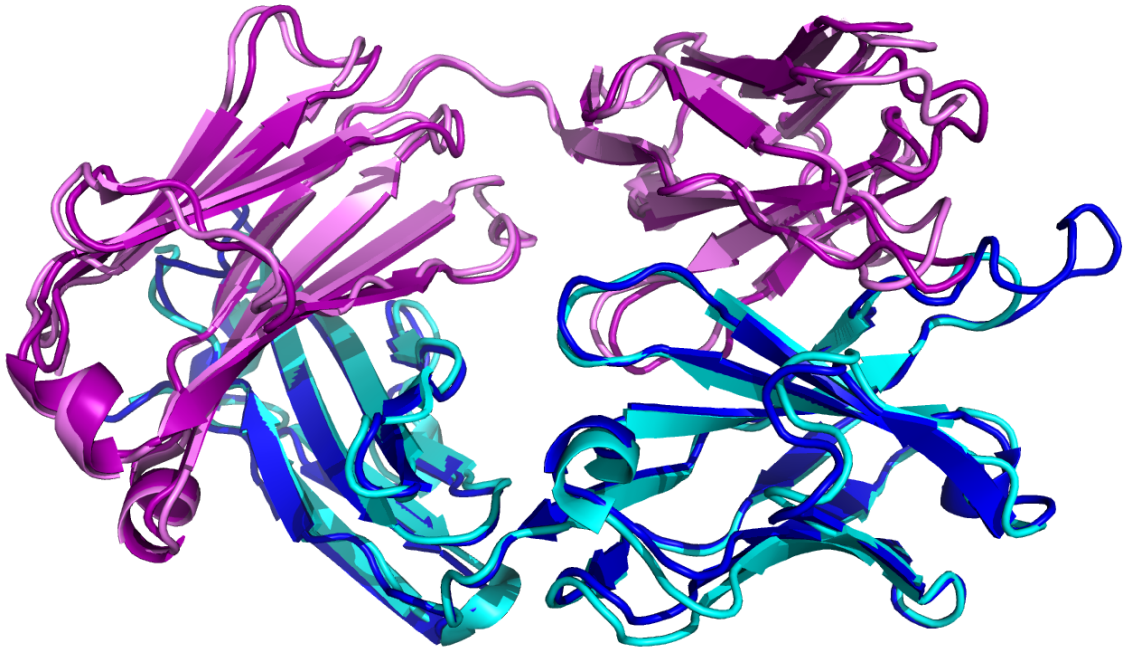


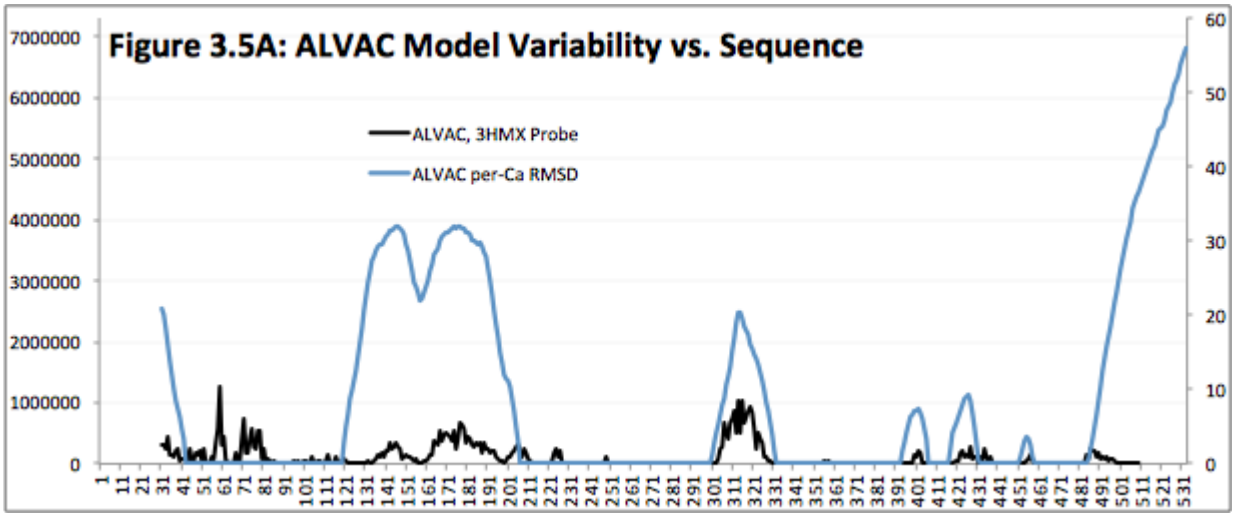
Figure 3.4B: Structural Alignment of the 3HMX and 2NY7 Fab Models Used as Probes. 2NY7 (darker blue and purple for heavy and light chains, respectively) has a comparatively long CDRH3 loop, which protrudes past that of the 3HMX structure (cyan and violet for heavy and light chains) on the right-hand side of the figure.

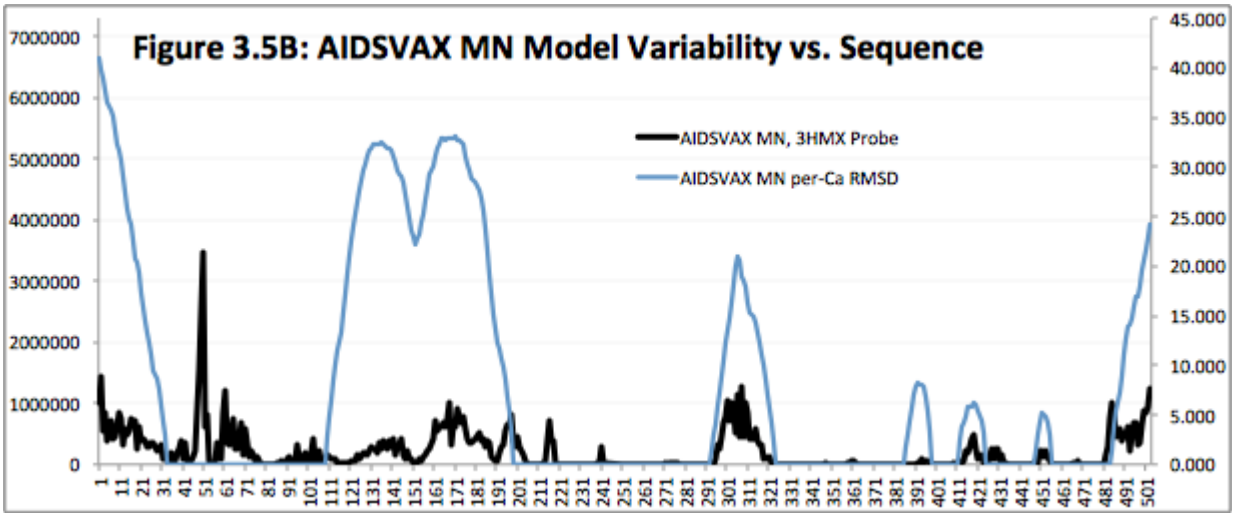
As the focus of the present work is on scoring the accessibility of orientations rather than finding Fab orientations with maximal possible interaction surface, we performed patch analysis on our gp120

conformational ensembles using the 3HMX model of the utsekinumab antibody identified by the b12 benchmark.

Rosetta Structural Prediction Generates Diverse Conformational Ensembles

Conformationally plausible, unranked ensembles for gp120 target sequences were generated using the Rosetta protein structure prediction package. In particular, the Remodel program was used to generate protein backbone conformations for known flexible regions absent or variable in extant gp120 crystal structures; these regions comprised the N and C termini of the protein, as well as the variable loops.





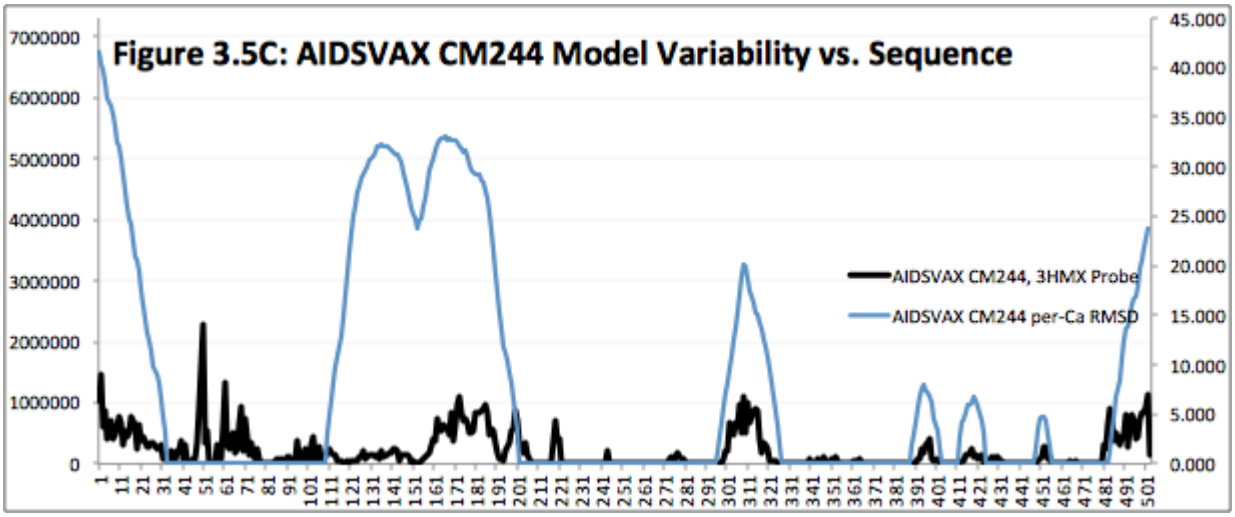


Figure 3.5: Model Variability vs. Amino Acid Sequence Position. Rosetta-built loops show variability in regions of sequence (blue) corresponding to variable gp120 loops identified in prior studies. Regions accessible to antibodies frequently map to these positions, as well as the N-terminal region; flexible 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 inter-structure alignment here.

Following protein model generation, N-linked glycosylation sites were modeled using high-mannose core structures; initial placement on sequence-identified glycosylation sites was followed by monte carlo steric clash minimization of inter-sugar bond rotations. A second round of protein backbone modeling was performed for the ALVAC model set, which was placed near the z=0 plane as a fully glycosylated, "finished" model and linked to a transmembrane helix model corresponding to the C-terminal sequence by a second round of fragment-assisted modeling in Remodel.

Glycosylation Sterically Hinders the Majority of Surface Epitopes

A central assumption underlying the EPIMAP method is that glycosylated models will show broadly attenuated binding to their protein component by an all-atom Fab model. To test this contention, we performed patch analysis of fully glycosylated and deglycosylated versions of the same set of 1000 gp120 models and compared patch counts across the two sets.

Comparison of the accessible surface patches between glycosylated and unglycosylated models reveals that, as expected, patches detected by the EPIMAP algorithm are often reduced on a per-residue basis by the presence of glycosylation; while certain areas of the glycoprotein escaped significant steric hindrance from nearby glycans, the majority of surface-exposed residues found on the deglycosylated

models were subject to some steric hindrance, and many residues in close proximity to glycosylation were entirely obscured. Regions of high exposure in the glycosylated model with relatively low glycan-mediated blocking mapped to the V2 and V3 variable loops (residues 160-200 and 300-330 respectively), both of which are targets of interest for anti-HIV broadly neutralizing antibodies.

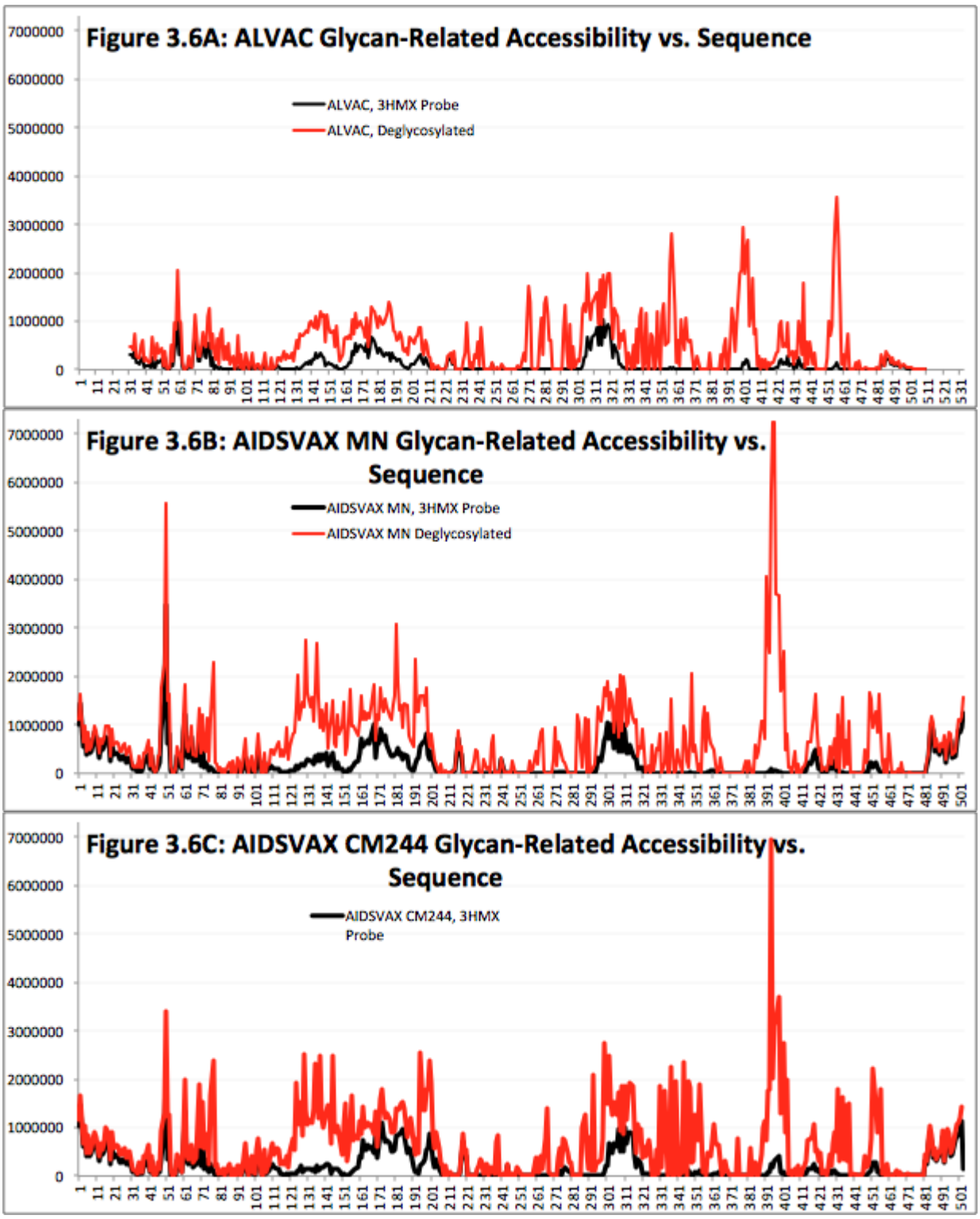
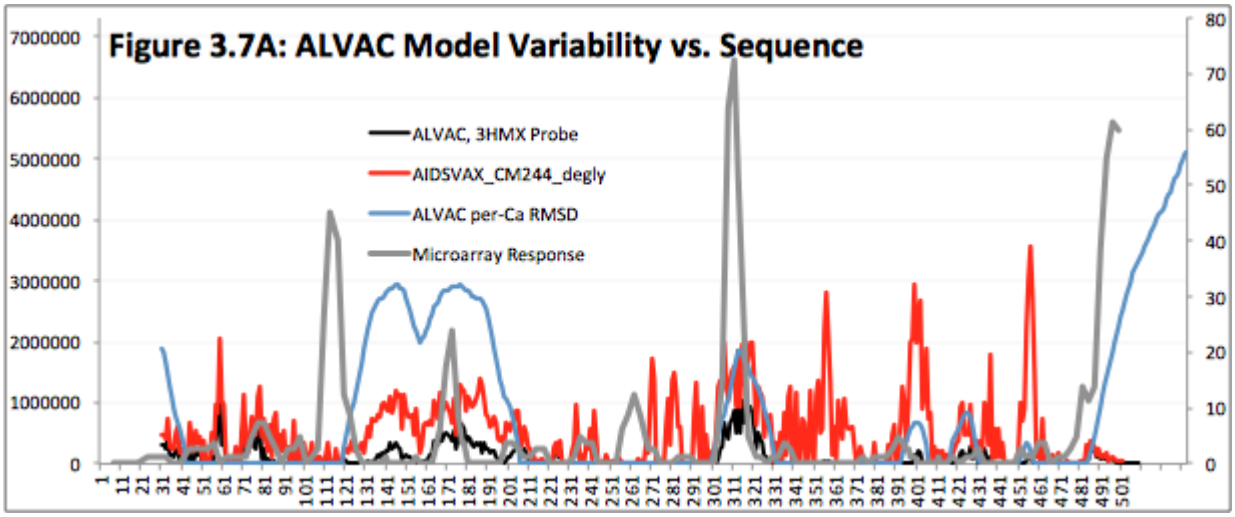
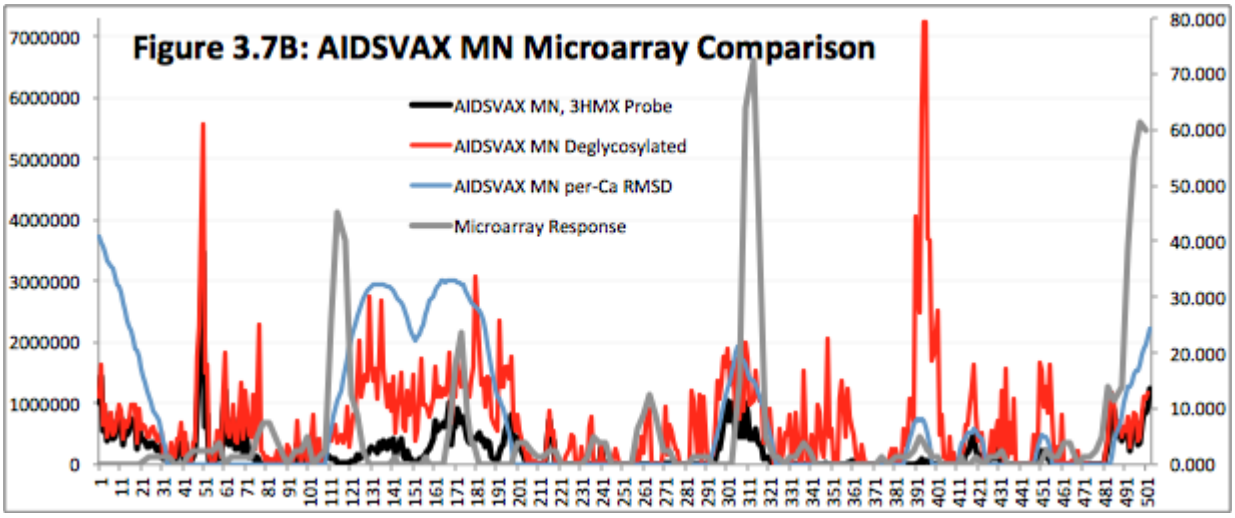


Figure 3.6: Glycosylation-Induced Reductions in Epitope Accessiblity. Deglycosylated models' patch count outputs (red lines) are uniformly higher than those of the fully glycosylated models (black lines); regions particularly well-shielded by glycans include the V1 loop (residues 120-15) and the V4 loop (residues 390-400).

EPIMAP Analysis of Computationally Modeled RV144 Components Broadly Matches Microarray Immunogenicity

Potential validation tools for the EPIMAP method would ideally provide experimental measurements of epitope accessibility. However, B cell epitope antigenicity imposes an additional, poorly predictable confounding factor when attempting to experimentally assay accessibility; the best available surrogate data within our capabilities is linear peptide immunogenicity for an overlapping peptide microarray spanning the vaccine components used in the RV144 trial.





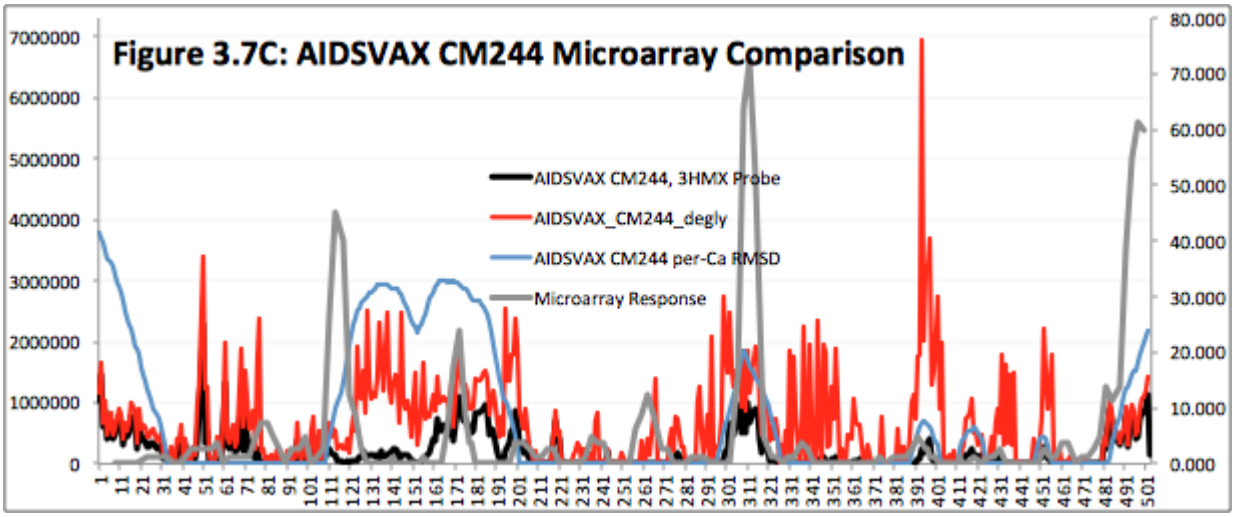


Figure 3.7: Microarray Experimental Evaluation of EPIMAP modeling. Microarray detection of 15mer peptides spanning the sequence of gp120 (gray) largely tracks with EPIMAP predictions (black); exceptions are in regions which are inflexible, as assayed by alpha carbon RMSD (blue) and covered by glycosylation (red exposure vs. black glycan-shielded responses).

The major discrepancy between EPIMAP-predicted accessible regions and observed antibody binding sites occurs in gp120 residues 110-120; this region, comprising a short stretch of the bridging sheet at the base of the V1/V2 loop[45], is known to be highly flexible in its native viral context. It seems likely that our conformational ensemble's limitations preclude our successful description of the bridging sheet's accessibility.

Of the three viral sequences assayed, the CM244 sequence demonstrates the best agreement between our accessibility predictions and microarray assay results. CM244 elicited the most antibody response of any of the three vaccine components[46]; thus, the possibility exists that decreased similarity between EPIMAP and microarray results for the anti-MN and -ALVAC responses is partially due to immunological factors outside the computational method's focus.

Comparison of antigenic regions detected by the microarray experiment and predicted accessible regions from EPIMAP reveals compatibility, but not ranked-order matching, between experimentally and computationally identified residues of interest on the surface of gp120 when considered as a whole. On the other hand, ranked-order matching for regions of well-modeled response - such as the V2 and V3 loops - 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.

Discussion

Given the pressing importance of developing even modestly effective methods for analyzing the immune response against heavily glycosylated viral entry 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.

EPIMAP results largely described the regions of high response to linear epitopes on gp120 as assayed by our collaborators' microarray experiments; the 110-120 region represented the sole major undetected antigenic region. The failure of EPIMAP to entirely identify all experimentally determined antigenic regions of gp120 most likely results from the limited sampling of protein conformations permitted by our initial model building, particularly in light of the data regarding glycan-mediated exposure and the boundaries of backbone flexibility selected for initial model building; the 110-120 sequence region lies at the outer edge of the extended V1/V2 flexible zone, increasing the modeled flexible length of which 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 available from 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 out of otherwise antigenic epitopes which are obscured for reasons of conformational flexibility 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 in 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 leave an opportunity 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 unglycosylated proteins[47] 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.

Materials and Methods

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[48], 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 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. Our conformational ensemble was based upon the following sequences:

ALVAC:
DNLWVTVYYGVPVWRDADTTLFCASDAKAQETEAHNVWATHACVPTDPNPQELHLENVTENFNMWKN
NMVEQMQEDVISLWDQSLKPCVKLTPLCVTLNCTNANVTNVKNITNVPNIIGNITDEVRNCSFNMTTELRD
KKQKVHALFYKLDIVPIEDNTSSSEYRLINCNTSVIKQACPKISFDPIPIHYCTPAGYAILKCNDKNFNGTGP
CKNVSSVQCTHGIKPVVSTQLLLNGSLAEEEIIIRSENLTNNAKTIIVHLNKSVEINCTRPSNNTRTSINIGPG
QVFYRTGDIIGDIRKAYCEINGTKWNEVLKKVTKKLKEHFNNKTIIFQPPSGGDLEITMHHFNCRGEFFYCN
TTRLFNNTCMENETMEGCNGTIILPCKIKQIINMWQGAGQAMYAPPISGRINCVSNITGILLTRDGGLNNTN
ETFRPGGGNIKDNWRSELYKYKVVQIEPLGIAPTRAKRRVVEREKRLFIMIVGGLVGLRIVFAVLSVVNRV
RQG

AIDSVAX_MN:
KYALADASLKMADPNRFRGKDLPVLDQLLEVPVWKEATTTLFCASDAKAYDTEVHNVWATQACVPTDPN
PQEVELVNVTENFNMWKNNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLNCTDLRNTTNTNNSTANNNS
NSEGTIKGGEMKNCSFNITTSIRDKMQKEYALLYKLDIVSIDNDSTSYRLISCNTSVITQACPKISFEPIPIHY
CAPAGFAILKCNDKKFSGKGSCKNVSTVQCTHGIRPVVSTQLLLNGSLAEEEVVIRSENFTDNAKTIIVHLN
ESVQINCTRPNYNKRKRIHIGPGRAFYTTKNIIGTIRQAHCNISRAKWNDTLRQIVSKLKEQFKNKTIVFNQS
SGGDPEIVMHSFNCGGEFFYCNTSPLFNSTWNGNNTWNNTTGSNNNITLQCKIKQIINMWQEVGKAMYA
PPIEGQIRCSSNITGLLLTRDGGKDTDTNDTEIFRPGGGDMRDNWRSELYKYKVVTIEPLGVAPTKAKRRV
VQREKR

AIDSVAX_CM244:
KYALADASLKMADPNRFRGKDLPVLDQLLEVPVWRDADTTLFCASDAKAHETEVHNVWATHACVPTDP
NPQEIDLENVTENFNMWKNNMVEQMQEDVISLWDQSLKPCVKLTPLCVTLHCTNANLTKANLTNVNNRT
NVSNIIGNITDEVRNCSFNMTTELRDKKQKVHALFYKLDIVPIEDNNDNSKYRLINCNTSVIKQACPKISFDPI
PIHYCTPAGYAILKCNDKNFNGTGPCKNVSSVQCTHGIKPVVSTQLLLNGSLAEEEIIIRSEDLTNNAKTIIV
HLNKSVVINCTRPSNNTRTSITIGPGQVFYRTGDIIGDIRKAYCEINGTEWNKALKQVTEKLKEHFNNKPIIF
QPPSGGDLEITMHHFNCRGEFFYCNTTRLFNNTCIANGTIEGCNGNITLPCKIKQIINMWQGAGQAMYAP
PISGTINCVSNITGILLTRDGGATNNTNNETFRPGGGNIKDNWRNELYKYKVVQIEPLGVAPTRAKRRVVE
REKR

Antibody Fab Model Generation

Antibody Fab crystal structures were downloaded from the Protein Databank; Fab models corresponding to appropriate heavy/light chain pairs were manually selected 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.

Rosetta-Based Conformational Ensemble Modeling

Protein sequences for all RV144 components were threaded onto the 3JWD crystal structure of gp120[49]. The Remodel package in Rosetta[50] was used for initial model building for areas corresponding to the V1/V2, V3, and V4 variable loops, as well as the N and C termini; 5000 models were generated, from which the lowest-scoring 100 were selected for further loop relaxation. 5 structural models corresponding to each of the 100 low-scoring loop models were used as the basis for glycan placement based on asparagine residues in appropriate N-X-(S/T) motifs.

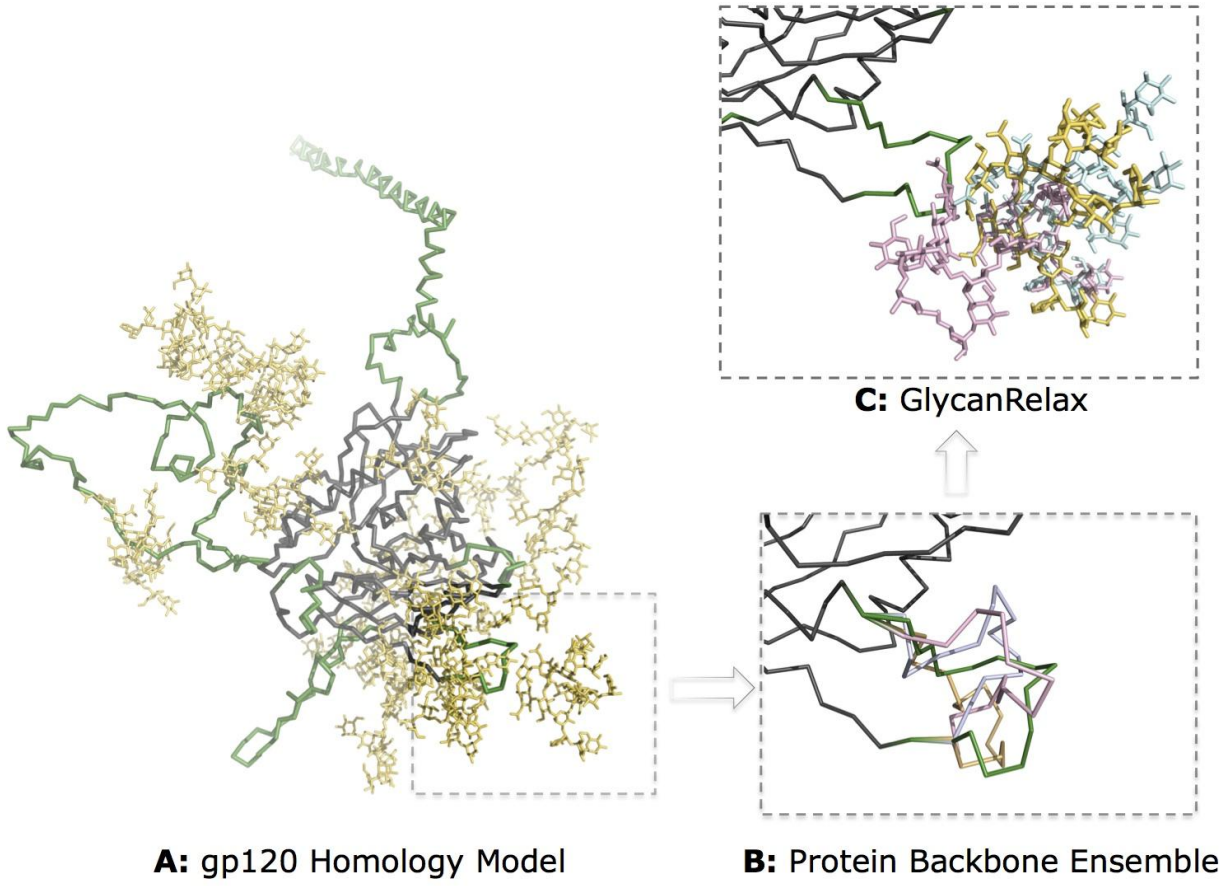


Figure 3.7: Rosetta Backbone Modeling and Glycan Relaxation. Regions of gp120 with known structural variability (green, panel A) are used to generate candidate loop conformations (panel B); the top 2% of gp120 conformations were used for subsequent glycan relaxation (panel C). Image courtesy of Sergey Menis.

Glycan Relaxation

Glycoprotein models were generated from protein input models via the placement of Man8 N-glycan core structures on glycosylation sites inferred from protein sequences. The Glycan Relax software used to generate models for EPIMAP prediction has been previously reported[5152]; 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.

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 - 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.

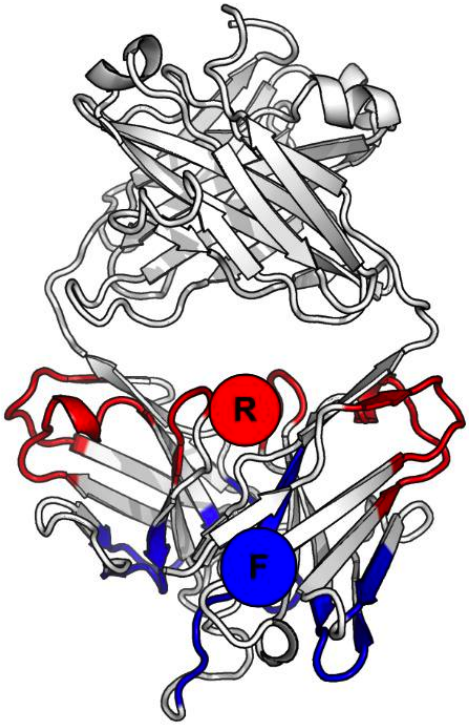


Figure 3.8: F_{AB} Model Probe Alignment Zones. Residues in the rear (R) and front (F) of the F_V are used to generate average positions; these form the vector which is used to align the probe to orientation vectors during patch prediction (Fig. 3.2A&B, center panels).

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 (Fig. 3.3), 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 executables 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.

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 Hz	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.

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 - 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, and Allan DeCamp and possibly Peter Gilbert as additional authors. Accordingly, while Chris Carrico assembled the manuscript above, much of the material underlying the methods - and in particular, prior work in B cell epitope prediction - was provided by Sergey and edited by Chris. Sergey will add additional material on Rosetta-based phases of the project and impose formatting for journal requirements to keep the shares of work on the manuscript even between both first authors.

Table 3.1: 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

References:

[33] Kulkarni-Kale U, Bhosle S, Kolaskar AS, *Nucleic Acids Res.* 2005. 33:W168-71.
[34] Andersen PH, Nielsen M, Lund O, *Protein Science* 2009. 15(11):2558-67.
[35] Sweredoski MJ, Baldi P, *Bioinformatics* 2008. 24(12):1459-60.
[36] Ponomarenko J et al, *BMC Bioinformatics* 2008 9:514.
[37] Yao B et al, *PLoS One* 2013. 8(4):e62249.
[38] Ackerman ME et al, *J. Clin. Invest.* 2013. 123(5):2183-92.
[39] Burton DR et al, *Cell Host and Microbe* 2012. 12(4):396-407.
[40] Schief WR et al, *Curr. Opin. HIV AIDS* 2009. 4(5):431-440.
[41] Plotkin SA, *Clin Infect Dis.* 2008. 47(3):401-9.
[42] Haynes BF et al, *NEJM* 2012. 366:1275-86.
[43] Zhou T et al, Nature 2007. 445(7129):732-7.
[44] Rolland M et al, *Nature* 2012. 490(7420):417-420.
[45] Kwong PD et al, *Nature* 2013. 393:648-59.
[46] Liu P et al, *JVI* 2013. 87(14):7828-36.
[47] Sampath S et al, *PLoS Pathogens* 2013. 9(6):e1003420.
[48] Alam SM et al, *J. Virol.* 2013. 87(3) p. 1554-68.
[49] Pancera M et al. *PNAS* 2010. 107(3):1166:71.

[50] Huang P-S et al. *PLoS One* 2011.
[51] Pancera M et al. *PNAS* 2010. 107(3):1166:71.
[52] Binley JM et al. *J. Virol.* 2010. 84(11):5637-55.

Chapter 4: Structural Analyses of Two-Loop Epitope Scaffolds

Introduction

Non-linear Epitopes Are Critical For Mediating Anti-HIV Serum Neutralization

Antibodies targeting an antigen of interest can, in general, bind to any exposed area of the target surface sufficiently unlike self epitopes to allow B cell proliferation. A common technique for mapping antibody epitopes on protein antigens is the presentation of an array of overlapping peptides spanning the antigen's sequence; thus, all surface-exposed residues are presented with at least a partial subset of their neighboring secondary structure, although such peptide "mimotopes"[53] typically bind more weakly than their parent antigens.

Immunodominant epitopes eliciting non-protective responses have long been identified as a key block in the development of a B cell-mediated vaccine against HIV[54]. Thus, peptide mimotope display - typically in the context of a suitable adjuvant to increase the peptide's immunogenicity and ensure T cell help for developing B cells - has been proposed, and tested at a preclinical level, as a vaccine strategy. While certain BNAb epitope regions on the *Env* surface are amenable to such a strategy, the majority of potently and broadly neutralizing anti-HIV antibodies discovered to date target complex epitopes comprising multiple stretches of protein sequence from non-adjacent regions of gp120 (Fig. 1.5, Table 1.1). Such epitopes require more complex methods of presentation; efforts to date have largely focused on shielding or excising immunodominant regions of gp120, but transplantation of desirable epitopes into scaffold proteins provides an alternative path forwards.

Two-Loop Grafting Poses Distinct Challenges

Much of the methodology behind the grafting of linear epitopes into heterologous proteins, as outlined in Chapters 2 and 3, depends upon the existence of a discoverable spatial arrangement of the linear epitope and its putative scaffold suitable for subsequent sequence optimization. This holds even for cases, such as the V1/V2 loop, where the insert of interest is of comparable size to the acceptor. Such arrangements occur as a natural consequence of the epitope graft matching procedure[55], in which a unique alignment oriented around an adjacent triplet of protein backbone heavy atoms serves to anchor one terminus of the epitope (Fig. 4.1A), leaving one and only one closure region for loop modeling to

accommodate; linear epitopes incorporating fixed secondary structure elements, such as the C-terminal alpha helix forming the gp41 MPER, offer a natural alignment point with a reasonable prediction of a final structure structurally similar to the design, even if some details of the modeled closure loop differ from the computational model.

The principal difficulty in designing epitope scaffolds for discontinuous epitopes arises from the increase in free epitope ends necessitating loop modeling; only one epitope subsection at a time may be used for aligning the acceptor protein, leaving 2N - 1 free loop ends requiring individual modeling for an N-loop epitope (Fig. 4.1B). As a low percentage of structures possess usable backbone topologies for any individual fixed structure, a vanishingly small number of scaffolds are topologically compatible with both epitope insertions, let alone capable of design into stable, well-folded protein scaffolds.

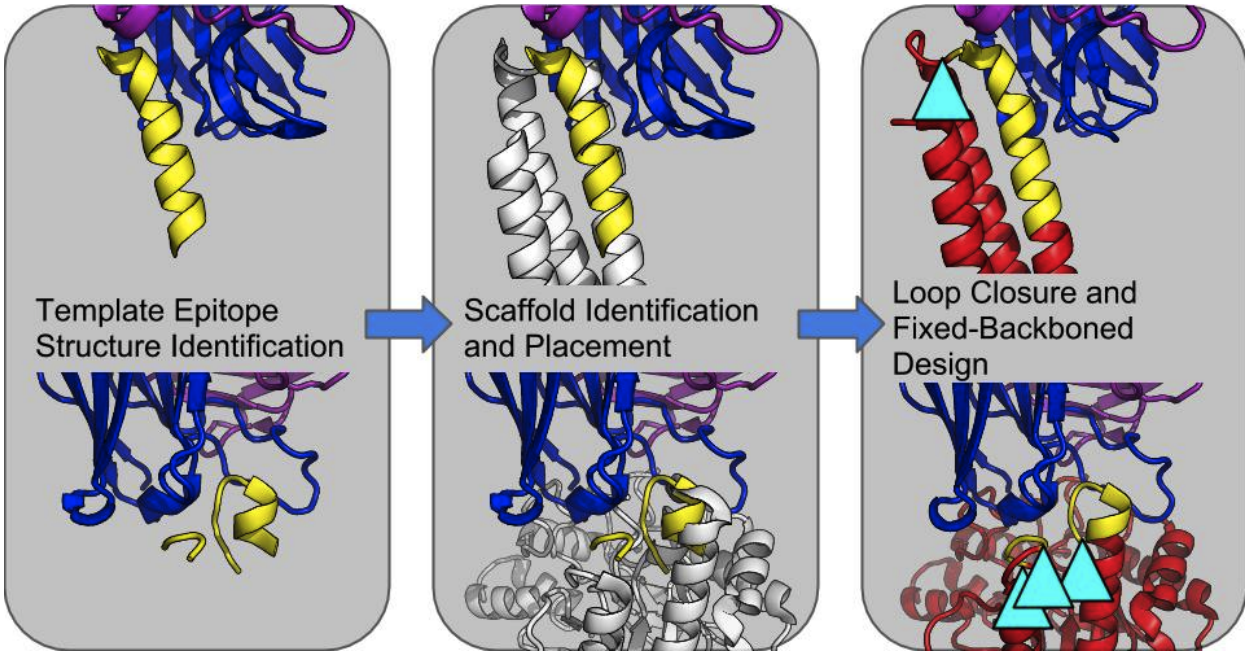


Figure 4.1: Epigraft vs. Multigraft Epitope Scaffold Design Methods. Both Epigraft (top) and Multigraft (bottom) use crystal structures of epitope:antibody complexes as structural templates (yellow, left panel) to identify compatible scaffold proteins (white, center panel) for subsequent, computationally expensive loop modeling and fixed-backbone computational design protocols. Forming the desired, high-affinity scaffolds (red with yellow epitope-bearing regions, right panel) requires modeling only one loop (cyan triangles, right panel) for Epigraft, but three for Multigraft as at most one loop may be used to generate the epitope:scaffold spatial arrangement.

Two-Loop Scaffolds Bind the b12 BNAb With High Affinity and Mimic gp120's Structure

Work led by Azoitei and Correia succeeded in developing a family of two-loop scaffolds exhibiting high affinities for the b12 broadly neutralizing antibody[56]. Initial computational design work yielded binding on the order of 100 uM, near the outer limits of detection by the surface plasmon resonance (SPR) method used to screen designs; two rounds of yeast surface display optimization using computationally selected library variation succeeded in driving scaffold:gp120 affinities to the 10 nM range, comparable to the native b12:gp120 interaction. These "2BODX" scaffolds derived from an endoglucanase from *Thermobifida fusca*, a highly thermostable enzyme with excellent stability and solubility.

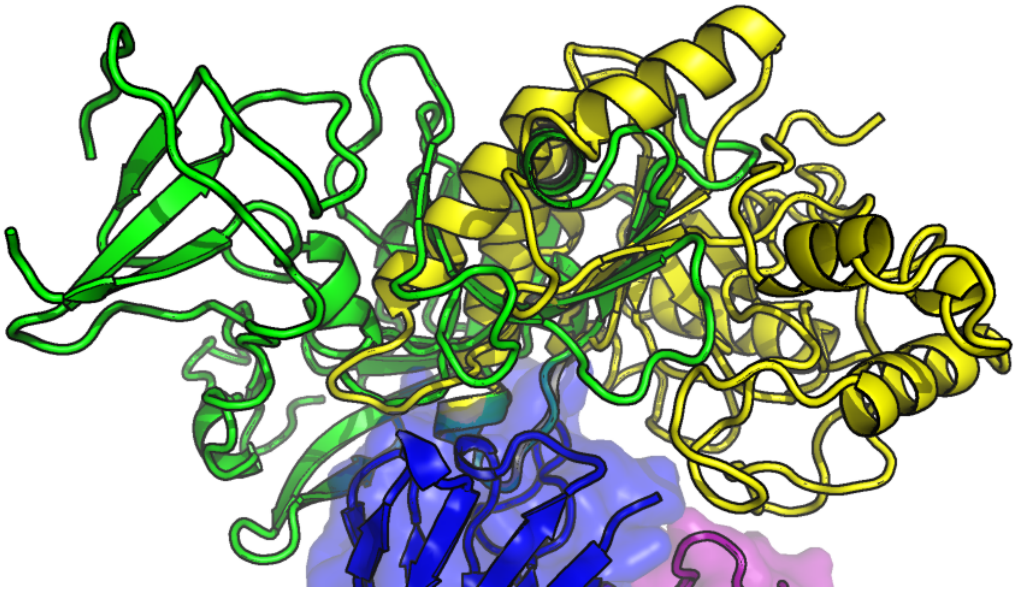


Figure 4.2B: Relative Arrangements of the 2NY7 gp120 and 3RU8 2BODX43 Complex Crystal Structures. The b12 Fab (blue heavy chain, purple light chain) as crystallized with 2BODX43 (yellow) is shown, with the truncated gp120 construct from the 2NY7 structure (green) placed by the 2NY7 Fab structure's binding region (not shown) with that of the 3RU8 Fab structure.

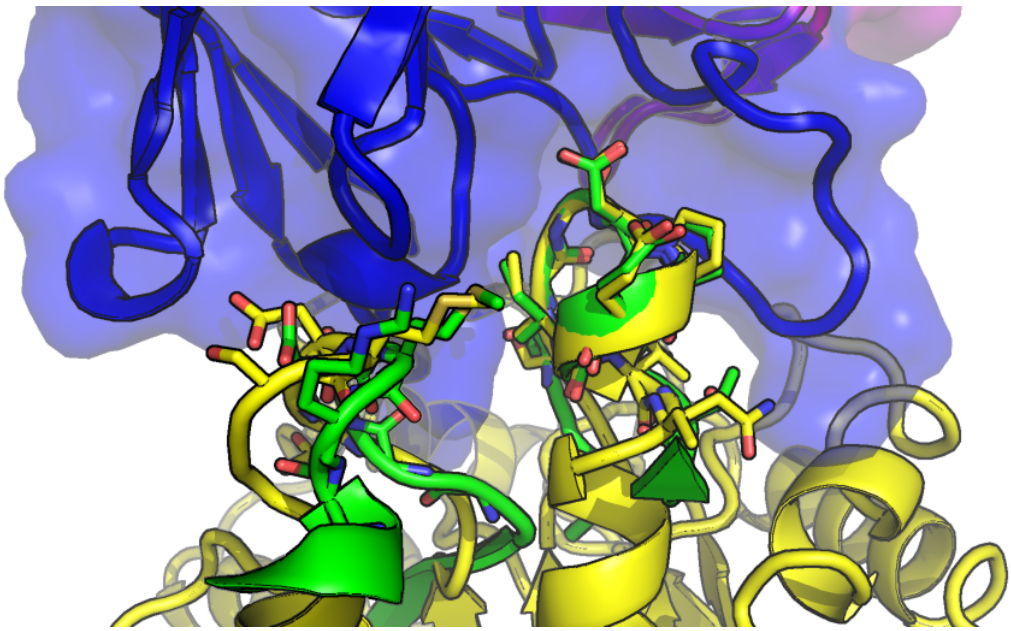


Figure 4.2C: Comparison of the gp120 and 2BODX43 Epitope-Bearing Loops. Contacts to the 2NY7 gp120 (green) and 3RU8 2BODX43 (yellow) are similar, and primarily mediated by the b12 Fab (blue, from 3RU8).

My work with these scaffolds, termed the 2BODX family from the PDB accession code for their template structure, began with the determination of the bound b12 Fab:2BODX43 complex. This structure, previously published with the initial design work, confirmed that the bound state of the scaffold's epitope-bearing regions is largely similar to that of the template region on gp120. As the project had been ongoing for some years at that point, and as the crystal structure of the b12:2BODX43 complex provided a complete story regarding the design process, work on examining the unbound structure of the scaffolds did not make the initial publication; those efforts, and ensuing work based on the first unbound 2BODX scaffold structures, is presented here.

Results

Crystal Structure of the 2BODX45 Unbound Scaffold

Of the three 2BODX scaffolds produced for the initial phase of crystallography attempts, 2BODX43 was the first to produce crystals of Fab:scaffold complex; pursuing these crystals for the complex crystal structure resulted in a depleted stock of unbound 2BODX43 insufficient for unbound structure determination. Of the two remaining scaffolds, 2BODX45 generated initial hits in 100 nL

crystallization conditions which successfully scaled up to provide large, rhomboid crystals with high-intensity diffraction yielding indexable spots past 1.5Å. Enforcing a signal intensity:noise cutoff of 3:1 for the outermost shell of diffraction spots led to a 1.52Å data set which yielded a structure largely matching the bound conformation (Fig. 4.3A).

Significant differences were largely restricted, unfortunately, to the epitope-bearing region (Fig. 4.3B); while a major portion of the intramolecular contacts characteristic of the bound loop conformation for residues 75-83 are present in the 2BODX45 structure, differences in backbone dihedrals at and immediately surrounding resides 77 and 82-85 allow the intervening epitope-bearing residues to shift away from the other half of the presented gp120-like surface. This structural rearrangement is likely assisted by extensive contacts formed between residues 79-84 and a neighboring copy of the protein in the crystal lattice (Fig. 4.3C).

A point of interest regarding the overall structure of the unbound 2BODX45 scaffold is the significantly higher degree of conformational flexibility exhibited by the unbound epitope as compared with the bound-state epitope from the b12 Fab:2BODX43 complex structure. 2BODX45's high-affinity binding to b12 requires that it sample a bound state similar to 2BODX43's; a flexible epitope region not only allows such sampling but also implies that the dissimilar conformation seen in the unbound structure may depend largely on crystal contacts rather than a stable, undesirable solution structure.

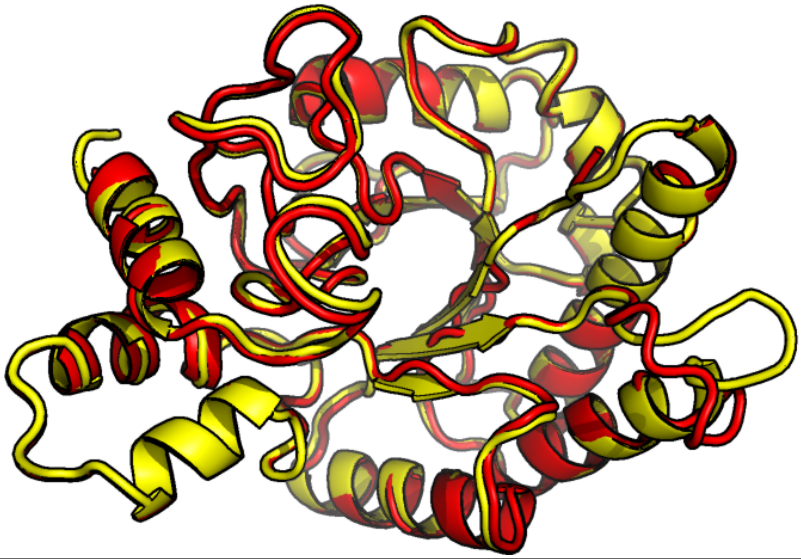


Figure 4.3A: Overall Structural Comparison of 2BODX Bound vs. Unbound Structures. 2BODX43 as bound to b12 (yellow) and 2BODX45 unbound (red) are structurally aligned, demonstrating the high degree of similarity outside of the epitope-bearing region.

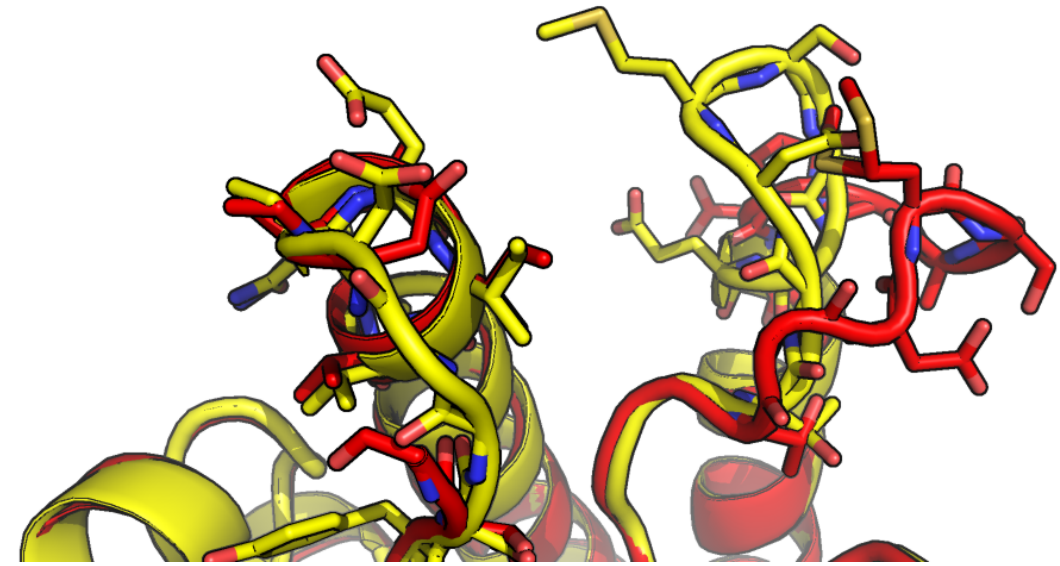


Figure 4.3B: The Unbound 2BODX45 Epitope-Bearing Region Differs from the 2BODX43 Bound State. Epitope-bearing loops corresponding to residues 111-115 (left) and 76-84 (right) are structurally distinct between the bound and unbound states of 2BODX scaffolds of largely equivalent affinities.

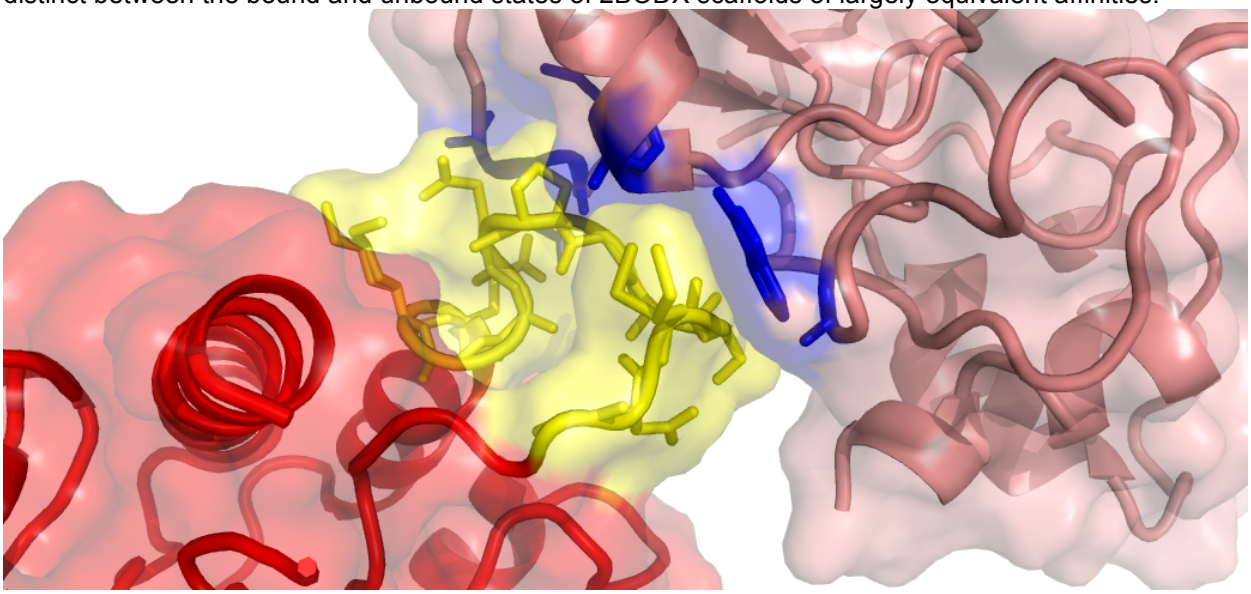


Figure 4.3C: 2BODX45 Residues 77-88 Contact a Neighbor in the Crystal Lattice. The epitope-bearing loop comprising 2BODX45 residues 77-88 (yellow) forms extensive contacts with residues 181-186, 210-211, and 234-237 (blue) on an adjacent copy in its crystal lattice.

Crystal Structure of the "Bound-Conformation-Selective" 2BODX49 Scaffold

In an attempt to crystallographically observe a bound-like conformation for an unbound scaffold, mutations were designed into the 2BODX45 sequence to disrupt the crystal contacts between the undesirable bound conformation and the aforementioned surface on its crystal lattice neighbor. Crystallization trials were attempted on three mutants; one, 2BODX49, yielded diffraction data at 1.35A

which enabled determination of its unbound structure. The conformation of the epitope-bearing loops in the unbound 2BODX49 structure is quite similar to that exhibited by unbound 2BODX45 (Fig. 4.4A).

The two discernible effects on the 2BODX49 structure which may potentially be attributed to the disruption of the previous crystal-packing interface. First, the electron density observed for the epitope-bearing loops reveals fewer structured residues in the 2BODX49 structure than in that of 2BODX45; second, a change in the crystal lattice between the two structures changes the area of 2BODX49 in contact with residues 77-88 on its packing neighbor (Fig. 4.4B). These changes suggest both that the effort to disrupt the particular packing arrangement observed in 2BODX45 was successful, and that the previously observed bound conformation - while perhaps favored by crystal contacts - does not require the particular contacts found in the 2BODX45 crystal structure. Thus, the 2BODX epitope's unbound conformational ensemble is likely distinguishable from its b12-bound state by a major contribution from the unbound loop conformation seen in the 2BODX45 and 2BODX49 structures in solution, implying possibly suboptimal binding affinity and epitope presentation.

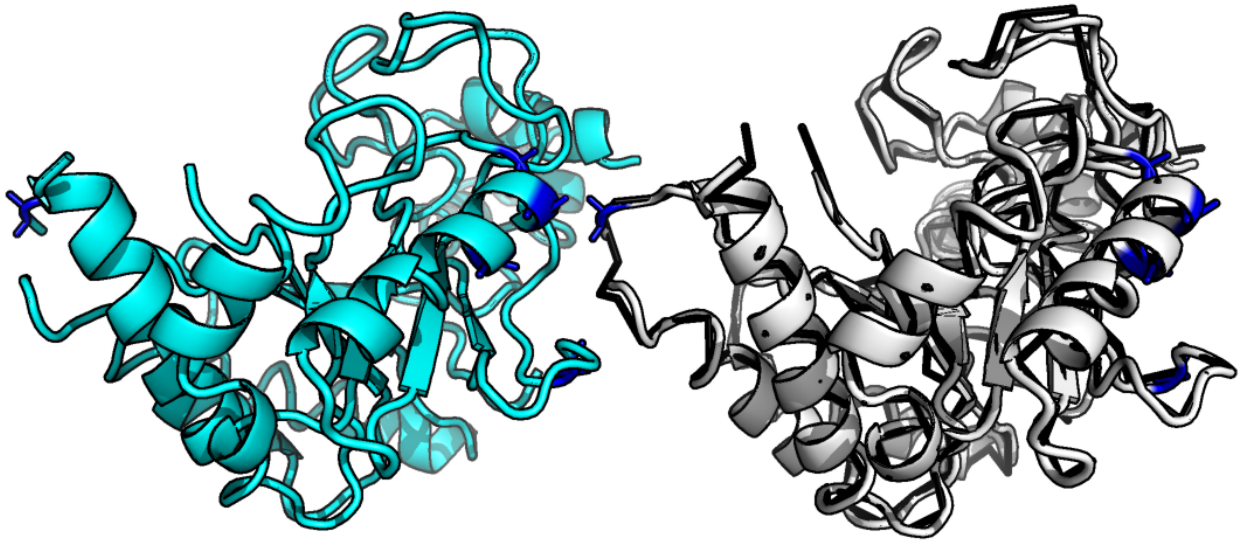


Figure 4.4A: 2BODX49 Crystal Contact Disrupting Residues Fail to Re-Create the Bound 2BODX43 Conformation. While the epitope regions of the unbound 2BODX49 structure (white cartoon) appear less defined than those of the 2BODX45 structure (black ribbon), mutating five residues (blue) between the two sequences does not succeed in preventing a largely similar contact surface with the neighboring subunit (cyan).

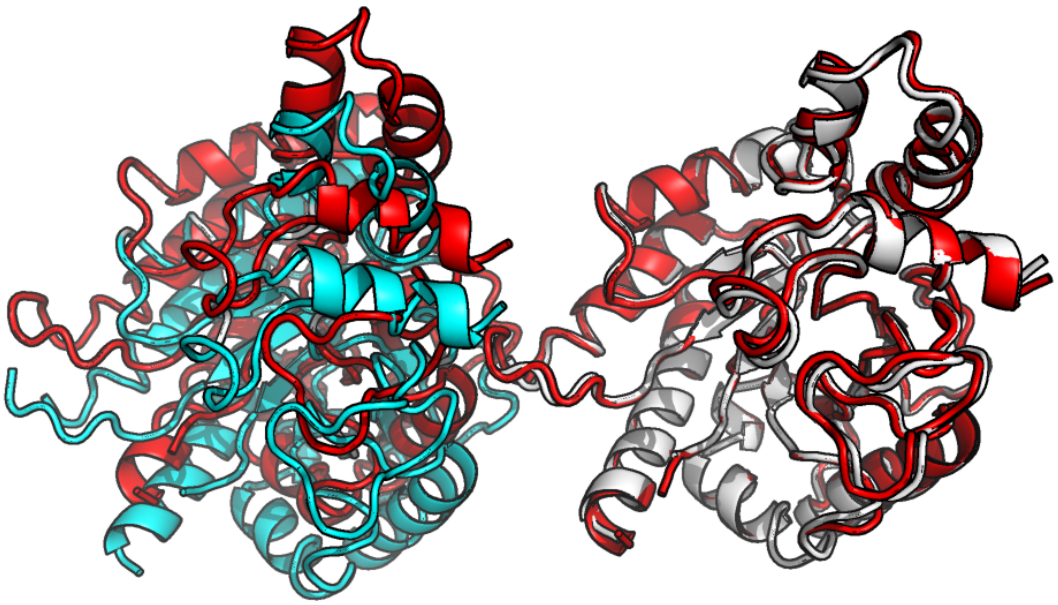


Figure 4.4B: 2BODX49 Crystal Packing Differs from 2BODX45. Alignment of the symmetry-related pairs of unbound scaffolds from the 2BODX45 (red) and 2BODX49 (white, cyan) crystal structures reveals a shift in lattice packing which corresponds to a change in contact between residues 77-88 and the neighboring lattice cell.

Crystal Structure of a Random-Library-Derived 2BODX Scaffold

In a parallel optimization project to the 2BODX46-49 development trajectory, Mihai Azoitei conducted NNK library mutagenesis from the original 100 uM design as identified by SPR, using libraries of equivalent complexity as determined by combinatorial sequence diversity. The 2BODX60 construct which resulted from this yeast display procedure binds b12 with ~100 nM affinity; attempts towards both unbound and scFv-bound crystallization of the scaffold resulted in a structure of an A118V point mutant of the 2BODX60 (2BODX60M) scaffold alone. The epitope-bearing loops of this unbound 2BODX60M structure reveal an unbound conformation much farther than 2BODX45 from the 2BODX43 bound state (Fig. 4.5A,B).

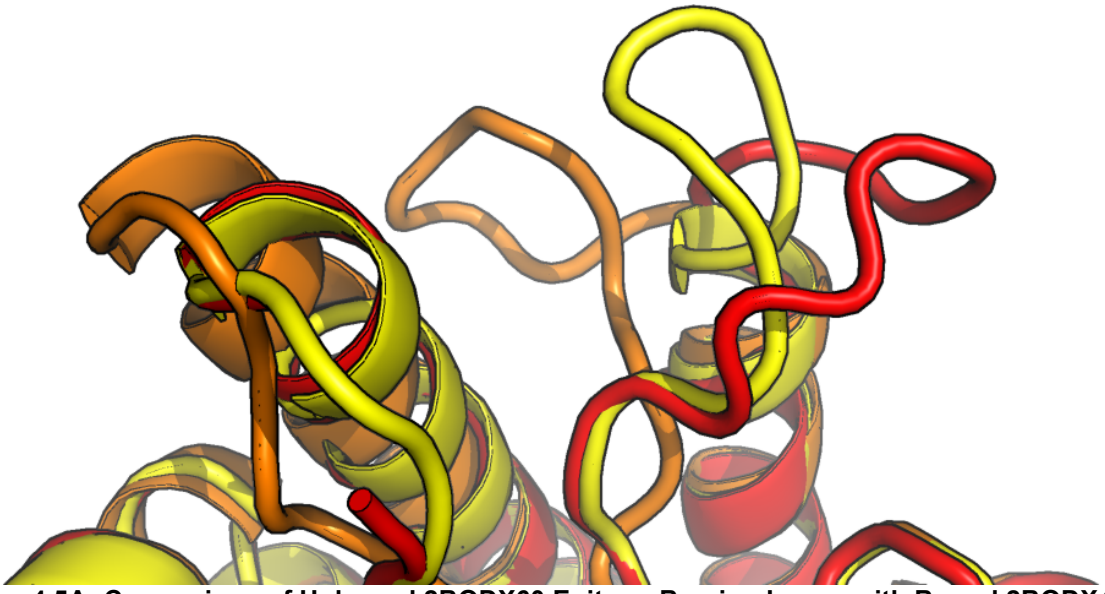


Figure 4.5A: Comparison of Unbound 2BODX60 Epitope-Bearing Loops with Bound 2BODX43 and Unbound 2BODX45. Residues 104-128 (left) and 72-85 (right) differ in conformation between the NNK-library-generated 2BODX60 (orange) unbound structure and both the bound 2BODX43 (yellow) and unbound 2BODX45 (red) crystal structures.

Figure 4.5B: 2BODX60M Epitope-Bearing Loops vs. 2BODX43 and 2BODX45, Tilted.

Comparison of Unbound 2BODX Scaffolds Reveals Superior Bound-Conformation Recapitulation via Computational Library Selection

While the affinities of the 2BODX60M and 2BODX43/45/48/49 scaffolds are within a factor of 30, the unbound 2BODX60M conformation bears considerably less resemblance than that of unbound

2BODX45 to the b12-bound conformations in either the Fab:2BODX43 or the scFv:2BODX48 structure. We can describe this observation in greater detail by comparing the backbone torsion angles and inter-residue bonds shown in the unbound structures with those seen in the bound states:

$$BBTorsionDifference(residue\ i) = |\Phi(i)' - \Phi(i)| + |\Psi(i)' - \Psi(i)| + |\Omega(i)' - \Omega(i)|$$

where backbone dihedral angles Φ(i)', Ψ(i)', and Ω(i)' correspond to the unbound crystal structure of interest and their counterparts Φ(i), Ψ(i), and Ω(i) to the bound 2BODX43 structure.

Assuming regular bond geometry, backbone dihedral angles form a complete description of a protein's fold and secondary structure; thus, the per-residue sums of differences in backbone dihedrals between two structures of equal length give us insight into where structural deviations occur (Fig. 4.6A, 4.6B). Investigation of the epitope-bearing loops and surroundings reveal large deviations for both scaffolds adjacent to the epitope-bearing loop containing residues 77-85, but the differences for 2BODX60M are more widespread, offering quantitative support for greater bound-state recapitulation by the unbound 2BODX45 structure.

Figure 4.6A: 2BODX45 Dihedral Differences Track to Residues 75-85. Backbone dihedral variations (white = 0°, red = 540°, linear scale) for residues 112 and 114 are induced by the absence of residue 113 in the crystallographically observed electron density. Colorization is proportional to the sum of backbone dihedral angle differences.

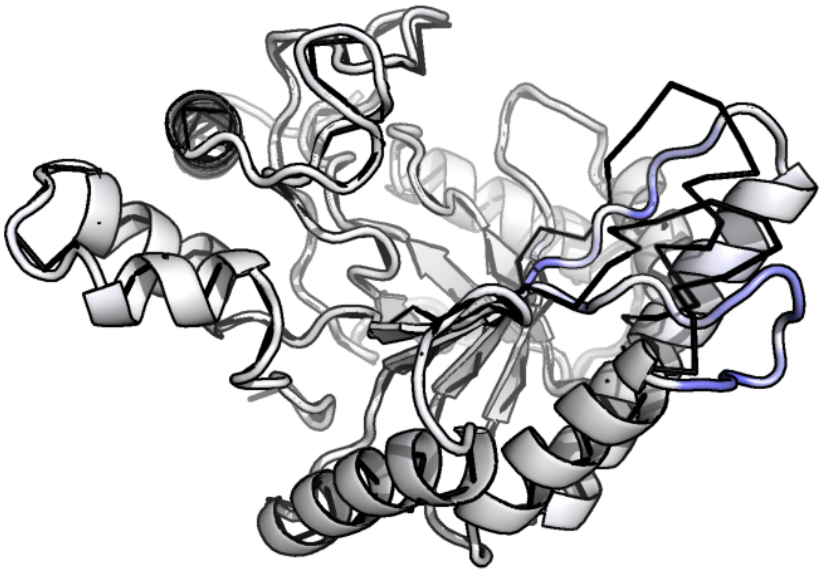


Figure 4.6B: 2BODX60M Dihedral Differences Are Found in Both Epitope-Bearing Loops. Residue ranges 73-85 and 107-120 both exhibit dihedral differences relative to 2BODX43. Colorization is proportional to the sum of backbone dihedral angle differences.

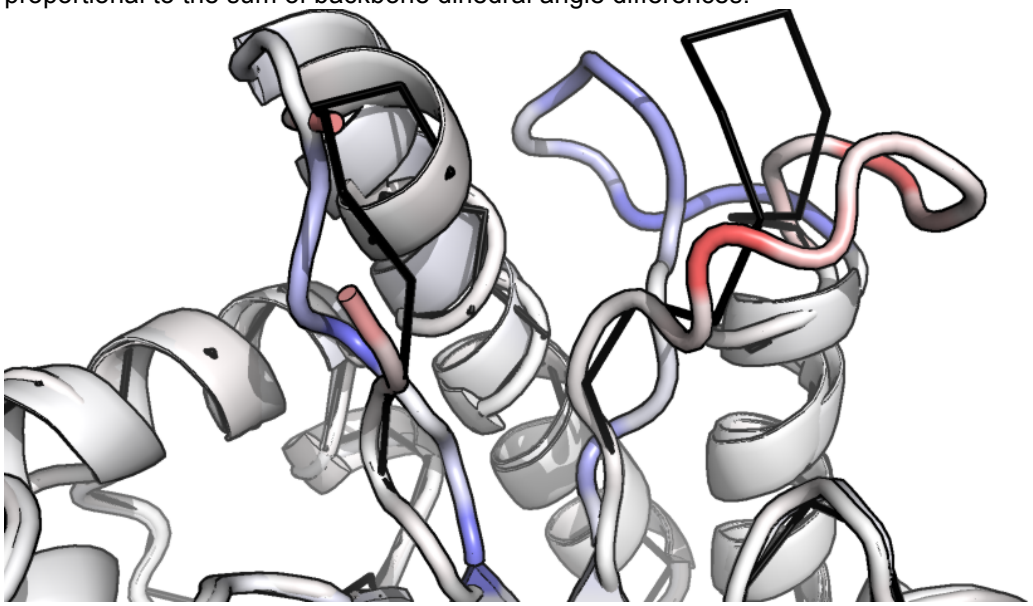


Figure 4.6C: Comparison of 2BODX45 and 2BODX60M Dihedral Differences. The epitope-bearing loops of 2BODX45 (white-red scale) and 2BODX60M (white-blue scale) both exhibit differences from 2BODX43 in residues 77-84, but 2BODX60M additionally differs in residues 73 and 107-120. Colorization is proportional to the sum of backbone dihedral angle differences.

Discussion

The discovery that the unbound 2BODX45 scaffold failed to recapitulate the epitope conformation of the bound 2BODX43:b12 Fab complex was unquestionably a disappointment; the failure of the 2BODX49 scaffold to rescue the bound conformation in the unbound state was perhaps unsurprising, but left the two-loop scaffold design project in an undesirable state. Given the demonstrably superior

recapitulation of structural elements of the bound 2BODX43 scaffold conformation by 2BODX45/49 as compared with 2BODX60M, however, we are left with a more heartening perspective on the success of two-loop scaffold design and a path for further methods development.

Binding affinity for the b12 antibody was used as the experimental sorting criterion for the development of the 2BODX scaffolds. However, the desired end product of epitope scaffolding is an immunogen capable of re-eliciting antibody activities comparable to the template epitope used for selection rather than a high-affinity antibody-binding reagent; the computationally optimized libraries appear to yield superior, though imperfect, structural recapitulation of the scaffold as originally computationally designed. Thus, we have evidence suggesting that computational optimization may enhance library selection as typically performed when the end product is characterized by structural similarity to a desired conformation in addition to more directly measurable properties such as affinity; computational design limits the disruptions to a desirable structure which are introduced in the selection process.

Materials and Methods

Protein Synthesis and Purification

2BODX scaffolds were expressed as commercially ordered, codon-optimized genes in the pET29b(+) vector in the Arctic Express strain of E coli. The 2BODX-containing vectors were transformed into bacteria by heat shock at 42 C for 30 seconds as per standard protocols; liquid cultures were performed using 5 mL overnight tubes and subsequent inoculation of multiple Erlenmeyer flasks containing 1L fluid volume each with 50 mg/L kanamycin. Expression was induced at an OD600 of 0.6 using 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG), and cells were pelleted and frozen after 4 hours of post-induction incubation at 37° C.

Bacterial pellets were resuspended in either Bugbuster solution (2BODX45 and 2BODX49) or a sonication buffer (50 mM Tris pH 8.0, 150 mM NaCl, 0.5% Triton X-100); all resuspensions were treated with the additions of 0.1 mM PMSF, 10 mg/L DNAse and 100 mg/L recombinant lysozyme. Resuspended pellets were lysed by 1 hour of incubation with gentle shaking at 25° C (2BODX45 and 2BODX49) or by cycles of sonication (10 minutes total sonication time, duty cycle 50%, 30s sonication intervals) on ice.

Lysed pellets were clarified by centrifugation and 0.22 um filtration, and the resulting supernatant was applied to equilibrated Ni-NTA resin.

Ni-NTA resin with bound lysate was collected in column format, resuspended in 10 mL of wash buffer (20 mM NaPO4 pH 7.4, 300 mM NaCl, 25 mM imidazole), and allowed to settle without fluid flow; the resulting packed columns were washed with 3 column volumes of wash buffer by gravity flow, and subsequently eluted by the addition of 2 column volumes of elution buffer (20 mM NaPO4 pH 7.4, 300 mM NaCl, 300 mM imidazole). Ni-NTA eluate was concentrated to 2 mL total volume and purified by size exclusion chromatography on Superdex 75 resin running in either HEPES-buffered saline (2BODX45 and 2BODX49: 20 mM HEPES pH 7.4, 150 mM NaCl) or PNEA (2BODX60M: 25 mM PIPES pH 7.4, 150 mM NaCl, 1 mM EDTA, 0.02% NaN_3).

Crystallization of 2BODX Scaffolds

Crystallization trials were initially performed in 100:100 nL well:protein solution sitting drop conditions (80 uL well solution reservoir) using a Mosquito robot for drop setting; promising conditions were scaled up for secondary screening in 1:1 uL well:protein hanging drop conditions (1 mL well solution reservoir).

2BODX45 crystallized in 0.1M Bis-Tris pH 5.5, 0.06M ammonium acetate, and 18% PEG-10,000, forming rhomboid crystals of up to 0.2 mm per side in approximately 24 hours' growth which yielded high-quality diffraction. Attempts at cryoprotection, however, uniformly led to visible crystal cracking, multiple diffracting lattices in the collected data, or loss of diffraction quality; thus, the data used for refinement were collected from an unprotected crystal with sufficiently well-defined ice rings to unambiguously discard overlapping reflections. 2BODX49 crystallized successfully in 0.75M succinate, pH 7.0; as with 2BODX45, attempts at cryoprotection uniformly led to crystal damage or reduced diffraction quality.

The 2BODX60M from which diffraction data were obtained grew at the initial 100+100 nL testing scale in the JCSG+ broad matrix screen condition A8 (0.2M NH_4HCO_2, 20% PEG-3350) from a stock solution at a 280-nm absorbance of 31; it was not reproducible in the larger-volume hanging drop format. As per 2BODX45 and 2BODX49, the crystal was not subjected to cryoprotection beyond that provided by the crystallization condition itself.

Diffraction Data Collection and Processing

Initial screening of crystals was conducted on a Rigaku x-ray source using Cu Kα radiation (λ = 1.542Å). Data sets used for model building and refinement were collected remotely via the Advanced Light Source using beamlines 5.0.1 and 5.0.2, with subsequent data processing performed in HKL2000 (2BODX45 and 2BODX49) and d*TREK (2BODX60M).

Table 1: Crystallographic Data Collection and Refinement Statistics.

	2BODX45	**2BODX49**	**2BODX60M**
Data Collection			
Space Group	P 1 21 1	P 1 21 1	P1
Z-value	1	1	2
Cell Dimensions			
a, b, c (Å)	45.27, 66.56, 47.92	43.30, 65.90, 45.45	42.21, 44.40, 66.36
α, β, γ (°)	90, 109.9, 90	90, 105.0, 90	82.81, 81.92, 76.16
Resolution (Å)	50-1.52 (1.55-1.52)	50-1.35 (1.37-1.35)	42.9-1.80 (1.86-1.80)
R_{merge}	0.051(0.098)	0.041 (0.378)	0.070 (0.248)
I/σ(I)	49.6 (37.3)	36.9 (4.0)	6.6 (2.3)
Completeness (%)	90.7 (99.2)	95.1 (81.4)	96.3 (93.1)
Redundancy (outer shell)	3.5 (3.4)	3.8 (3.4)	1.92 (1.93)
X-Ray Source	ALS Beamline 5.0.1	ALS Beamline 5.0.1	ALS Beamline 5.0.2
X-Ray Wavelength (Å)	1.000	1.000	1.000
Refinenment			
Resolution (Å)	23.5-1.52	33.0-1.35	42.9-1.80
No. Reflections	35392	48739	39207
R_{work}/R_{free}	0.166/0.182	0.187/0.201	0.224/0.246
# Atoms			
Protein	2128	2012	3964
Water	436	223	195
B-factors, TLS Residual			
Protein	4.29		17.05
Water	15.21		24.02
Backbone Dihedrals			
Favorable (%)	97.76	98.48	97.93
Outliers (%)	0.00	0.00	0.00
MolProbity Score/Percentile	1.22 (97th)	0.96 (99th)	0.92 (100th)
RMS Deviations			
Bond Lengths (Å)	0.005	0.005	0.005
Bond Angles (°)	0.86	0.86	0.94
Estimated Coordinate Error, by Maximum Likelihood (Å)	0.039	0.035	0.106

Crystal Structure Solution, Model Building, and Refinement

Data for all structures was phased through the use of molecular replacement with the original 2BODX43 model as determined in the 3RU8 crystal structure used as a search model, with loops corresponding to mutated residues and adjacent regions missing. Model building and refinement were

carried out in Coot and Refmac5, respectively, as implemented in the CCP4 software package; an electron density contour level of +1.0σ was used to place residues. Water oxygens were placed for all structures only after complete placement of protein residues in all visible areas of unused density. TLS refinement was performed following placement and refinement of all possible protein and water atoms.

Unbound Structure Analysis and Comparison

Backbone dihedral angles for the 2BODX45, 2BODX49, and 2BODX60M structures were calculated using the web server implementation of VADAR. All figures were generated using MacPyMol, and data from VADAR angle analyses were transferred to PyMol colorizations using a Python script to set per-residue occupancy values.

Acknowledgements

As mentioned in the chapter introduction, the 2BODX scaffold family originates from computational work by Bruno Correia in tandem with experimental yeast display selection by Mihai Azoitei; the subsequent 2BODX60 selection project was conducted by Mihai Azoitei. 2BODX45 and 2BODX49 were expressed and purified by Alex Schroeter; I expressed and purified 2BODX60M.

Crystallization condition trials and optimization for 2BODX45 and 2BODX49 were conducted with assistance and advice from Meg Holmes, Matthew Clifton and Colin Correnti, who also assisted with crystal screening on the Cu Kα source at FHCRC. Matthew Clifton and Colin Correnti assisted and advised with remote data collection at the ALS, and - with, again, help from Meg Holmes - with data processing and refinement techniques, though I performed all computational operations leading to the finished structures as presented here. I performed all crystallization, data collection and processing, and refinement steps for 2BODX60M.

References:

[53] Lenstra JA et al, *J Immunol Methods* 1992. 152(2):149-57
[54] Robinson WE Jr. et al, *PNAS* 1990. 87(8):3185-9.
[55] Correia BE et al, *Structure* 2010. 18(9):1116-26.
[56] Azoiței ML et al, *Science* 2011. 334(6054):373-6.

Chapter 5: Biophysical Comparison of the b12 Broadly Neutralizing Antibody to its Germline Precursor

Introduction

Affinity Maturation Is Responsible For High-Affinity BNAb Binding

As discussed in Chapter 1, clinically isolated broadly neutralizing antibodies are products of an iterative process of mutation generating a polyclonal B cell response to an introduced antigen. As such, the germline precursor sequences from which the mature, high-affinity sequences derive typically exhibit low affinities for eventual target antigens; indeed, binding by certain anti-HIV antibodies' germline precursors - including b12, our current subject - is undetectable[57]. A major question for attempts at recapitulating clinically observed broadly neutralizing antibody responses is, then, whether proposed antigens can properly stimulate germline precursor antibodies of interest to develop desired antibody specificities. B cells displaying non-binding germline precursors could potentially be stimulated by heterologous antigens to give rise to an HIV-binding, partially mature BCR sequence; this has been proposed as a mechanism explaining the development of anti-HIV BNAbs.

As affinity maturation requires at least modest affinity for antigen by a germline antibody, mutational gain of affinity through the maturation process is likely to be gradual, and should typically add side-chain-mediated contacts or reinforce a desirable binding conformation in the maturing antibody's variable fragment (F_V). This paradigm suggests mechanisms for both surface-exposed and buried amino acid side chains as potential sources of affinity gain; as B cell selection is a stochastic process, mutations without discernible effects on binding may also be introduced during the maturation process. These mutations may have BCR selection benefits at the level of mRNA codon usage, protein folding, or stability; they may even be mildly disadvantageous. B cell maturation, like other selection processes, carries no guarantee of optimality in its results.

Canonical Protein:Protein Interface Theory Predicts "Hot Spot" Contacts

As used in the context of inter-protein interactions, "hot spot" residues are amino acids responsible for key thermodynamic contributions to a protein:protein interface. Hot spots are commonly observed in the course of mutational studies on protein:protein interfaces, particularly through the use of

alanine scanning[58]; as alanine possesses no side chain beyond a single beta carbon, it ablates side-chain-dependent interactions while leaving secondary structure relatively unaltered.

Mutational studies of antibody:antigen interactions have revealed amino acid positions of key biophysical importance to binding[59]. Such studies often yield antibody variants with improved antigen binding by finding residues detrimental to the antibody:antigen interaction which may be subsequently mutated away for the production of an optimized, recombinant antibody. Additionally, mutational studies of antibody binding offer insights for antibody elicitation by identifying desirable aspects of a clinically or biochemically observed monoclonal antibody whch may be used to optimize immunogenic presentation of its target antigen. Given the difficulties to date in successfully re-eliciting protective antibody responses against HIV, several previous studies have been attempted in this manner.

The b12 Broadly Neutralizing Antibody is a Well-Studied Test Case

First discovered in 1991[60], b12 binds to an epitope on the surface of gp120 which largely covers the functionally necessary CD4 binding site (CD4bs). The discovery of the b12 antibody, extracted from a phage-display library derived from peripheral blood mononuclear cells donated by an HIV-infected patient, revealed the possibility that a single antibody of particular binding specificity could neutralize a significant subset of infectious HIV strains; though both the breadth and potency of b12 neutralization have been surpassed by other anti-HIV antibodies, a wealth of work performed on the antibody during its long stint as the only known anti-CD4bs BNAb makes it an unusually fruitful target for germline progenitor analysis.

Past data of particular importance for the interpretation of include x-ray crystal structures of the intact b12 IgG[61], the F_{AB} construct of b12 bound to a truncated gp120 construct[62], and scanning mutagenesis studies focusing on the b12 CDRs identifying residues of key thermodynamic importance for the b12:gp120 interaction[63]. Based on these studies, past efforts have succeeded in creating both heterologous, non-gp120 proteins displaying the b12-bound epitope with comparable affinity to the template interaction[64] and an improved variant of the b12 antibody (3B3)[65] with higher affinity for gp120 and neutralizing potency against HIV; b12's interactions with gp120 are, therefore, as well-understood as those of any anti-HIV BNAb.

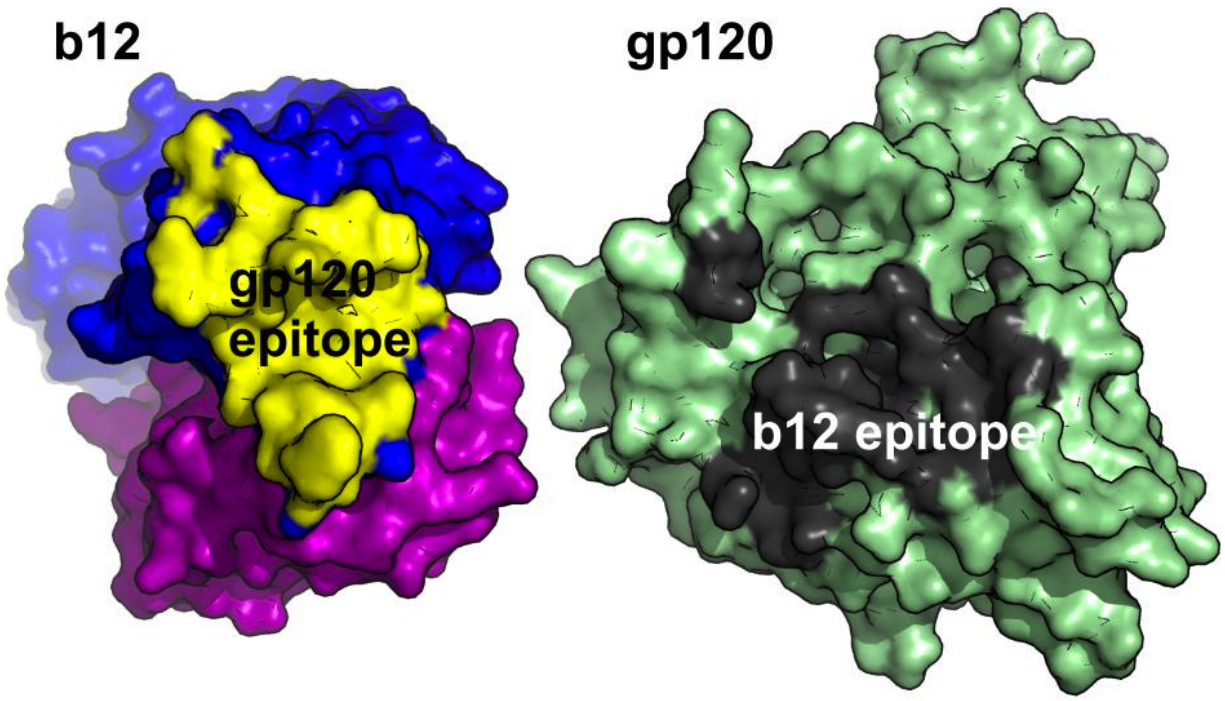


Figure 5.1: The b12 Epitope on gp120, and The gp120-Binding Paratope on b12.
As characterized in the b12 F_{AB}:gp120 crystal structure, the paratope on b12 responsible for mediating binding to the antibody's epitope on gp120 consists solely of heavy-chain residues (Fig. 5.1). According to the results of the past mutagenesis studies mentioned above, heavy-chain residues accordingly fill the role of interaction hot spots in the b12:gp120 interface; past attempts to re-elicit b12-like affinities have thus been guided by the attempt to engage these key residues with a candidate immunogen.

The sequence of the germline progenitor genes from which the b12 antibody developed was recently determined by PBMC deep sequencing from blood sample from the original donor. Sequence comparison of the b12 mature and germline sequences reveals 45 mutations in the F_V region of b12, distributed relatively evenly between the heavy (24 mutations) and light (21 mutations) chains (Fig. 5.2). Residues previously identified as interaction hotspots by CDR scanning mutagenesis studies number among the mutations identified; this opens the possibility of identifying a small subset of mutations acquired through b12's process of somatic hypermutation which, if successfully elicited by an immunogen, could yield a b12-like antibody with desirable neutralizing potency and breadth against HIV.

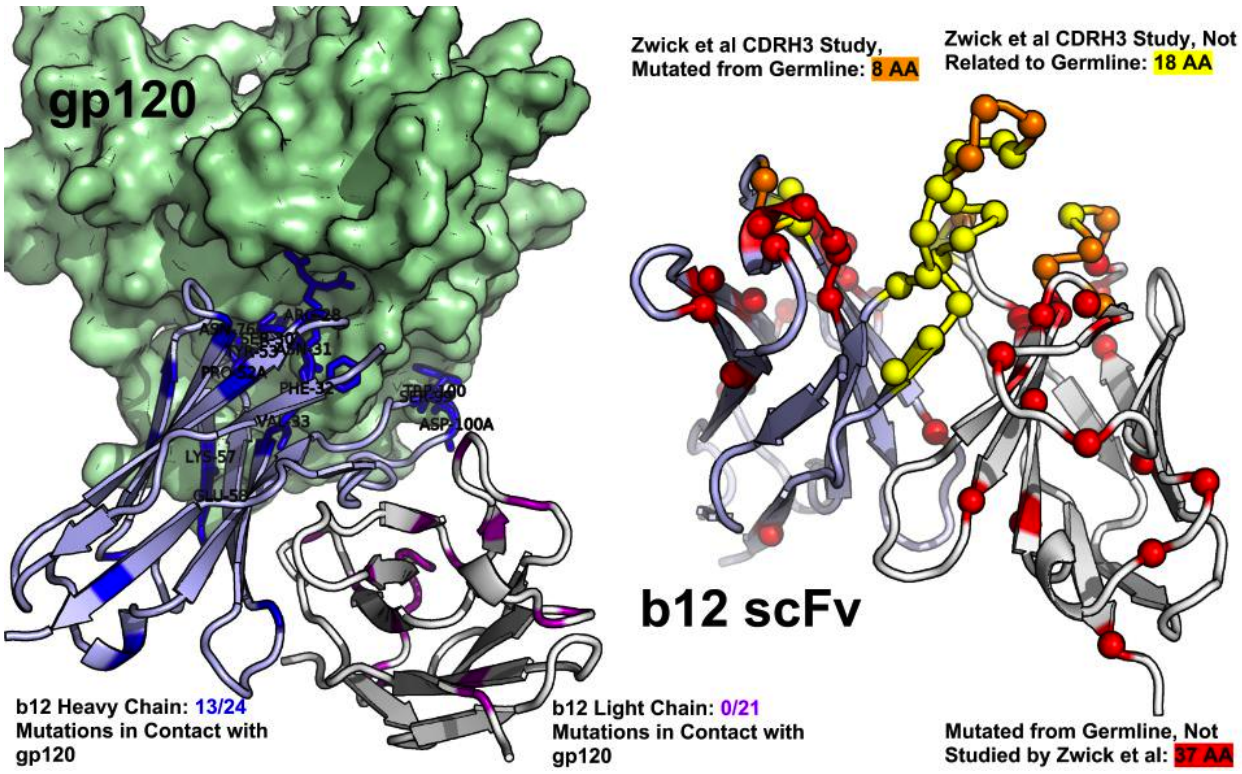


Figure 5.2: Mutations Between Germline and Mature b12. Left panel: gp120 atoms are within 6Å of 13/24 mutations in the b12 heavy chain, as seen in the 2NY7 crystal structure[62], but none of the 21 light chain mutations. Accordingly, prior mutagenesis study of b12[63] has focused on CDR residues (yellow and orange, right panel) known to be in contact with either gp120 or the residues mediating direct contacts, while leaving most mutations from the germline sequence (orange and red) unstudied.

Results

Point Reversion Mutant scFvs of b12 Are Expressible and Bind HIV gp120

Past studies on the b12 antibody have typically used intact IgG. While this form of the antibody is the most clinically relevant, its bivalent binding surfaces complicate the process of extracting usable information regarding the kinetics of binding. Some prior work has used the monovalent Fab construct, prepared by proteolytic cleavage of IgG, as a surrogate for the antibody to improve the quality of biophysical measurement. As efficient production of Fab requires IgG production followed by proteolysis, we produced point mutants for each of the 45 potential reversions from the mature b12 sequence towards its germline precursor as cleavable single-chain variable fragments (scFvs).

We determined the bulk dissociation constant of wild-type b12 scFv for two surfaces displaying one of two sequences of chemically coupled gp120, and found general agreement with literature values for b12 Fab:gp120 affinity for both gp120 sequences used (Fig. 5.3A&B). Screening of the 45 point

mutants with wild-type control runs at a concentration equaling the wild-type KD was performed against a second SPR chip prepared as per the first; all 45 mutant scFvs bound detectably to both the SF162 and QH0692 gp120 surfaces (Fig. 5.3C&D). Viewed in the structural context of the published 2NY7 crystal structure, the majority of mutations with measurable impacts on bulk affinity localize to the gp120-binding interface of the scFv; many, however, do not, and indeed one reversion with measurably decreased binding is located on the interface-opposite face of the scFv light chain, which lacks discernible interactions with gp120 according to the crystal structure (Fig. 5.3E&F).

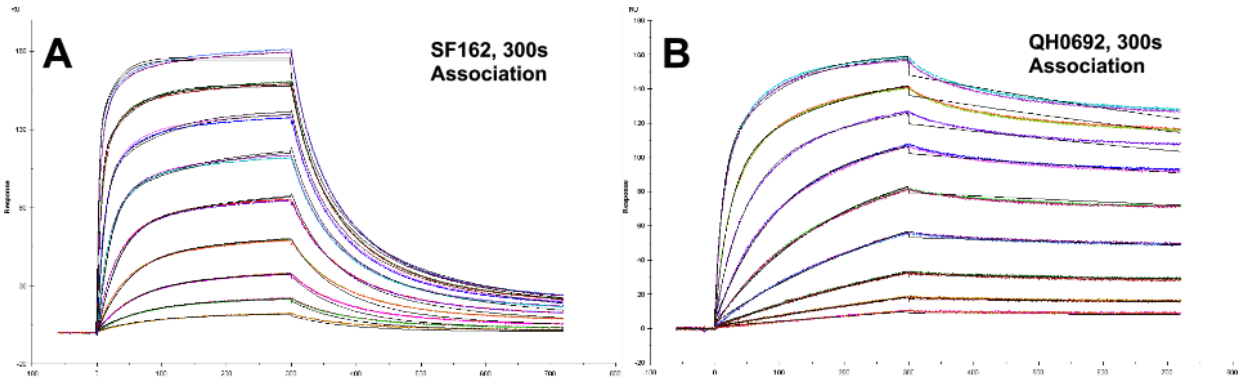


Figure 5.3A and 5.3B: b12 scFv 300s Association SPR vs. Amine-Coupled gp120 Surface. Single-chain Fv in twofold dilutions from 300 to 1.17 nM was used as a flow analyte over SF162 and gp120 surfaces; [scFv] for 50% maximal response was ~30 nM for both surfaces.

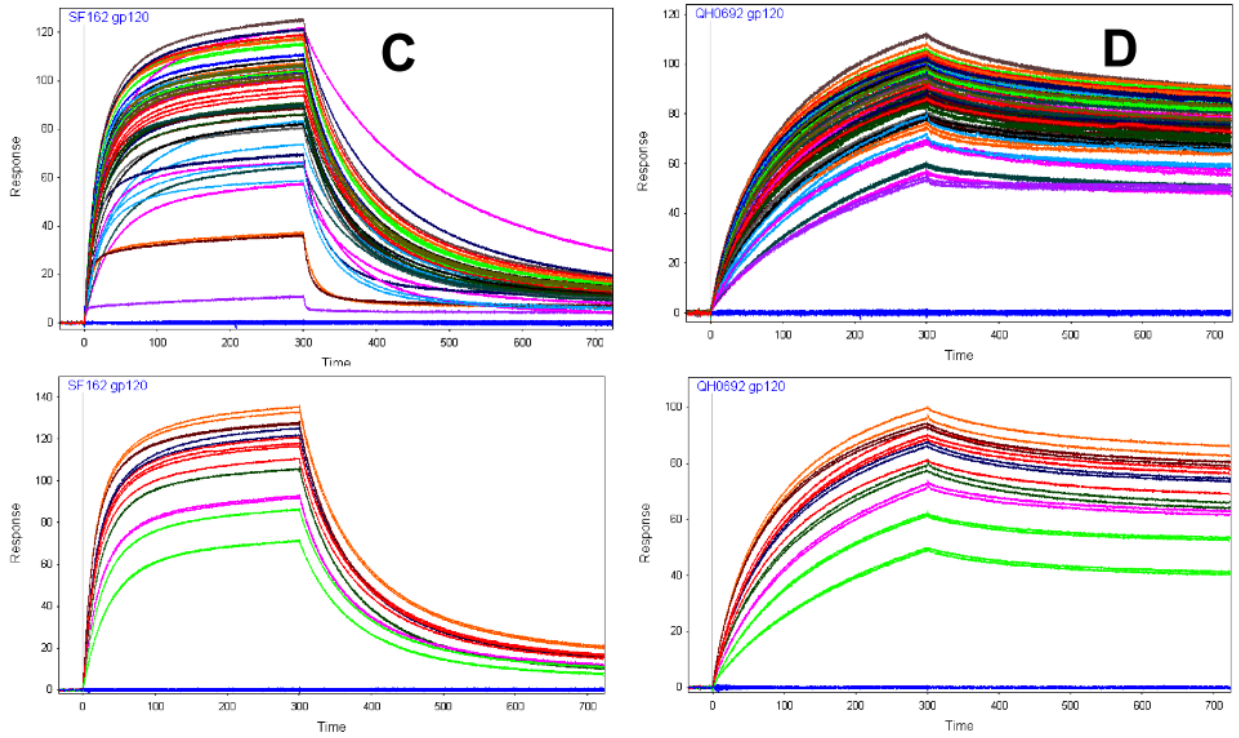


Figure 5.3C and 5.3D: b12 scFv Mutants 300s Association SPR. 38/45 point mutants in the top

panels, and the remaining 7/45 in the bottom, were used as flow analytes at 30 nM over SF162 **(C)** and QH0692 **(D)** surfaces; all mutants bound. Data corresponding to the t=300s response units are in Table 5.1.

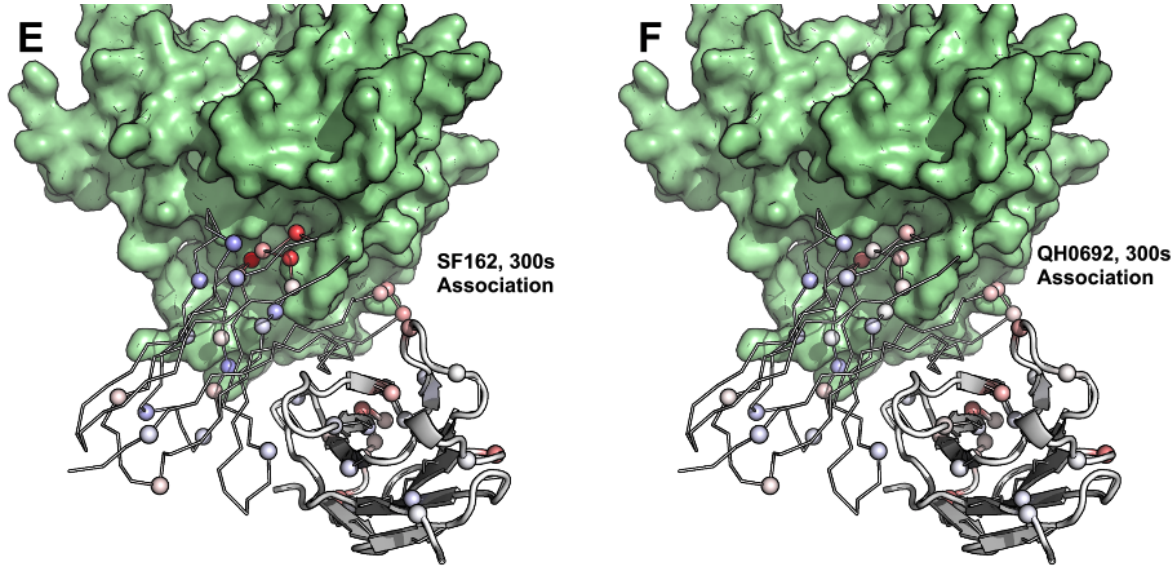


Figure 5.3E and 5.3F: Per-residue binding responses by b12 mutants at 30 nM are displayed using a red-white scale for ratios 0.0-1.0, and a white-blue scale for ratios 1.0-2.0. Note that both poor binders (red) and binders with higher response than mature b12 (blue) are more pronounced in SF162, particularly in the vicinity of the gp120 interface (identified as per figure 5.2).

Short Time Scale b12:gp120 Binding Improves Kinetic Parameter Fits

Binding of b12 to the bulk gp120 surface with sufficient time to reach an approximately steady state fits poorly to a 1:1 binding model; in particular, binding to the gp120 surface maintains a steady and approximately linear increase after an initial deceleration in binding rate, implying at least two gp120 species on the SPR chip surface with varying competencies to bind b12 (Fig 5.4A).

At short timescales, binding to surface component species with faster binding kinetics should dominate the SPR response measurement; we can thus achieve a reasonably good fit to this faster component's association and dissociation kinetics by shortening the association phase of the SPR experiment. Data from mutant scFv binding using a 150s association rather than the 300s used for the previous, equilibrium-phase experiment reveals significantly tighter binding by both the mutant scFvs and the wild-type controls (Fig. 5.4A&B), with calculated affinities of 11 nM for SF162 and 4.6 nM for QH0692. Structural context reveals a tendency for germline reversion mutations to both increase and lower b12's association rate (Fig. 5.4C&D), while these mutations typically increase the dissociation rate (Fig. 5.4E&F) of the b12:gp120 complex.

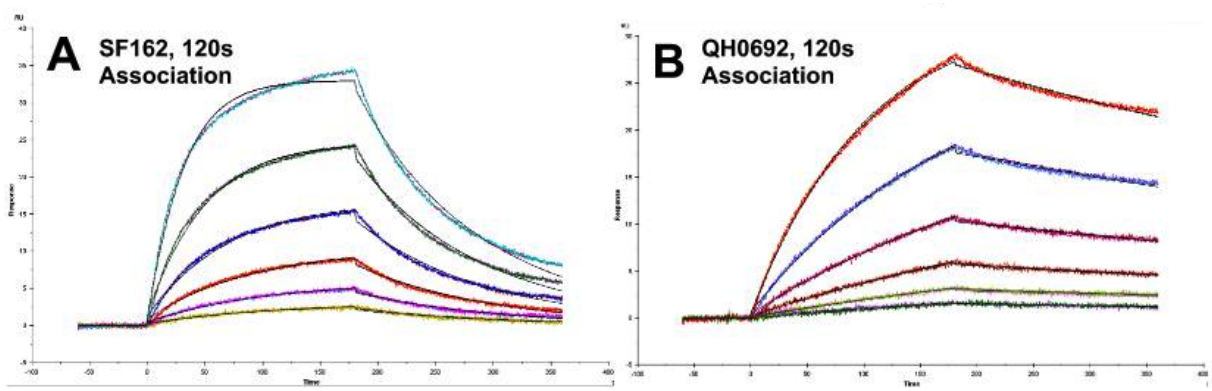
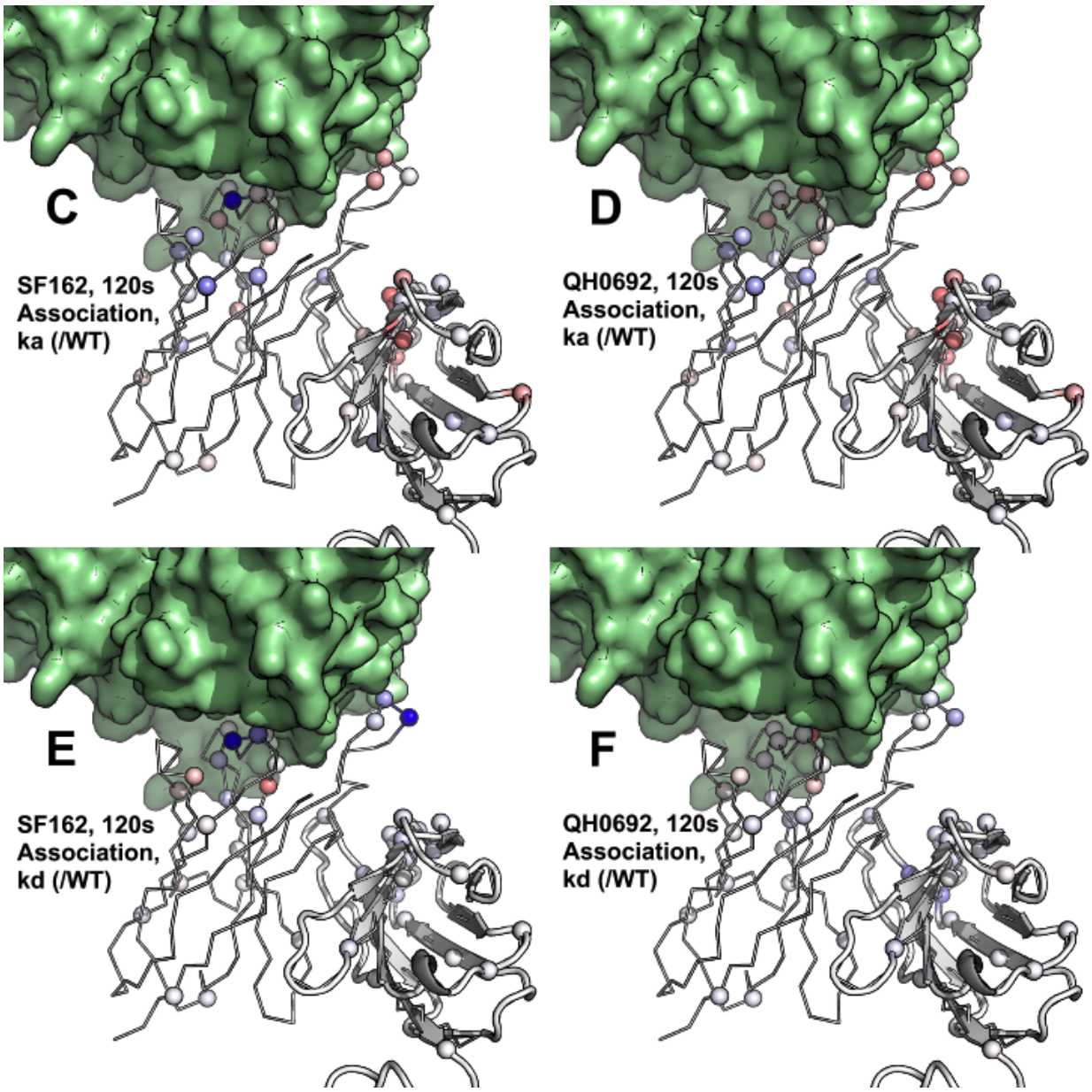


Figure 5.4: 120s-Association Biacore and Resulting Kinetic Parameters. Data series for SF162 **(A)** and QH0692 **(B)** gp120 sequences yield improved fits (black lines) for the 30 nM-0.94 nM data series pictured. Ratios of calculated association rates and dissociation rates for point reversion mutants vs. the mature scFv are displayed for SF162 in panels **(C)** and **(D)**, and for QH0692 in panels **(E)** and **(F)**, respectively; kinetic parameters for mutants are displayed using a red-white scale for ratios 0.0-1.0, and a white-blue scale for ratios 1.0-2.0. Thus, blue represents a faster on-rate than mature b12 in panels **(C)** and **(D)**, and a faster off-rate in panels **(E)** and **(F)**.



Thermal Stability of b12 Point Mutant Reversions

As many mutations acquired during b12's affinity maturation process map to side chains involved in hydrophobic core packing, we investigated potential effects on the thermodynamic stability of the b12 scF_V as assayed by a fluorescence thermal shift assay. A hydrophobic dye's intercalation into the partially unfolded scFv conformation yields an increase in dye fluorescence at or immediately above the protein's melting temperature (T_M); while direct determination of thermodynamic melting properties from this fluorescence assay is impossible, the effects of point mutations on T_M may be established by comparing the temperatures of peak fluorescence.

Individual mutations had relatively modest effects on b12's stability, with maximal shifts of 5 K. As might be expected from general principles of protein structure such as folding cooperativity, destabilizing point reversions outnumbered stabilizing point reversions (Fig. 5.5A), and both the most stabilizing and the most destabilizing reversions involved side chains buried in the scFv's core. Interestingly, four stabilizing reversions - mutations for which the mature b12 amino acid is less stable - localized to the C-terminus of the CDRH2 loop (Fig. 5.5B), implying some key role for this loop offsetting the loss of thermodynamic stability.

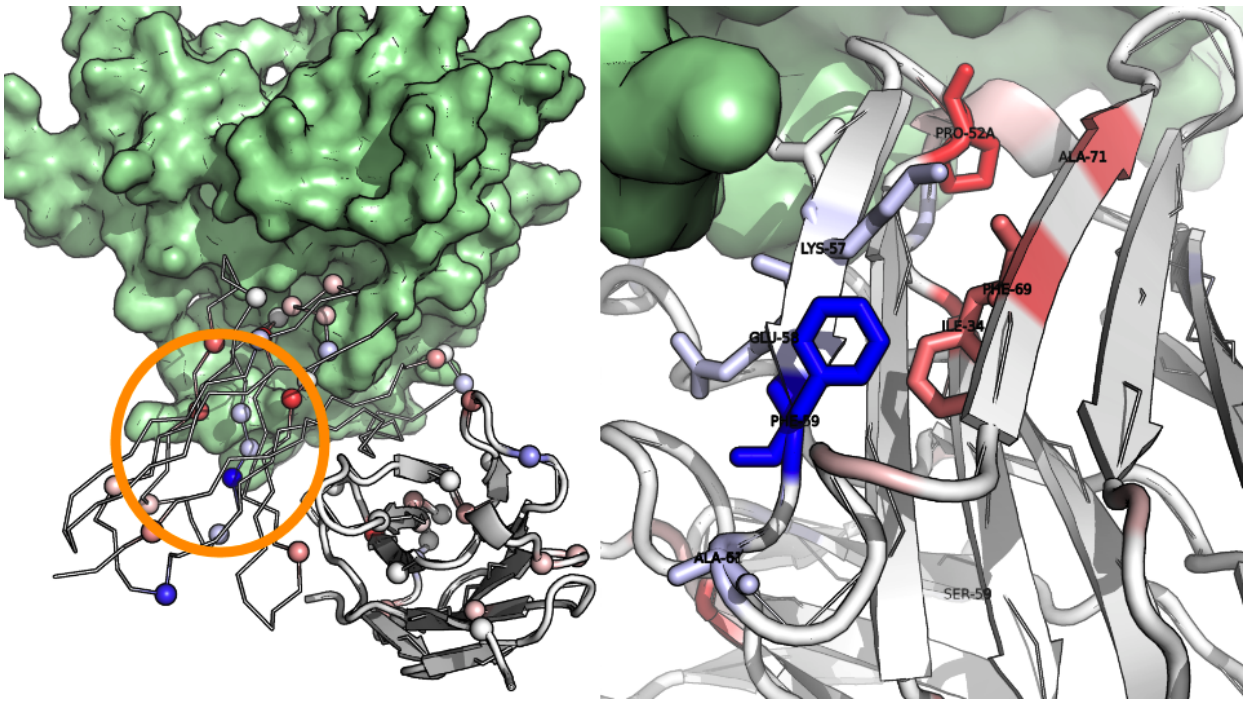


Figure 5.5: Melting Temperature Shifts from Point Reversion Mutations. Reversion mutants with melting temperatures ranging linearly from 5K lower (red) to 5K higher (blue) than mature b12 are distributed throughout the scFv structure **(A)**; one point of interest is the presence of a series of thermodynamically disadvantageous mutations acquired by b12 at residues 57-59 and 61 in the heavy chain **(B)**.

Comparison of Germline Reversion Mutants with CDR Mutants

Prior work on b12 by Michael Zwick and co-authors[63] included a large panel of b12 point mutants, many overlapping with the set described here. We are thus interested in both comparing any overlaps in the sets of mutants, and in comparing the overall performance of germline reversion mutants with the CDR-focused mutation set intended to investigate possible gains in affinity relative to the starting sequence. This prior work used competition ELISA and b12 point mutant F_{AB} constructs expressed in *E. coli*, with triplicate 10 mL overnight cultures pooled and used without subsequent purification of the

resulting bacterial lysate for many mutants; as such, variability between measurements is larger than that observed in our data. A particularly likely source of discrepancies is the authors' note that the Y53G mutant failed to purify well by acid elution from Protein G. As Zwick et al report b12 mutant affinities as fractions of mature b12 response against two gp120 sequences mutually exclusive with our test set of SF162 and QH0692, we will compare the average response of their sequences to each of ours.

Table 5.1: Collected Data and Comparisons for Germline Reversion Mutants

	SF162:			QH0692:			Zwick:	
Chain: Mutation	**Bulk RU/WT**	**ka/WT**	**kd/WT**	**Bulk RU/WT**	**ka/WT**	**kd/WT**	**KD/WT**	**TM Shift**
L:F21L	1.126	1.14	1.00	1.06	1.15	1.02		-1
L:S25A	1.049	0.96	1.05	1.00	0.94	1.13		0
L:H27Q	0.595	0.52	1.20	0.53	0.51	1.48		-1.5
L:I28aV	0.884	0.90	1.15	0.87	0.87	1.17		1
L:R29S	0.871	1.09	1.17	0.91	1.01	1.13	0	-1
L:R31S	0.887	1.02	1.14	0.92	0.95	1.16	0	0
L:R32Y	0.567	0.45	1.06	0.59	0.44	1.10	0	-1.5
L:V33L	1.121	1.14	0.95	1.07	1.17	0.94		-0.5
L:H38Q	1.041	0.95	1.07	0.95	0.96	1.10		0
L:V47L	0.779	0.57	1.00	0.77	0.60	1.00		0
L:H49Y	1.030	1.11	1.15	1.05	1.09	1.08		1
L:V51A	1.051	1.15	1.11	1.06	1.13	1.06		-1.5
L:N53S	1.078	1.07	1.05	1.03	1.08	1.04		0
L:S56T	0.779	0.74	1.04	0.76	0.75	1.07		-2
L:S59P	1.009	1.03	1.01	0.98	1.03	0.98		2
L:T76S	0.648	0.72	1.00	0.62	0.74	1.02		-1
L:V78L	1.029	1.07	1.02	1.03	1.07	1.02		-1
L:L85V	1.097	1.12	1.03	1.05	1.13	1.04		0
L:V90Q	0.898	0.85	1.14	0.87	0.88	1.39		-4
L:A93S	1.116	1.24	1.06	1.08	1.23	1.12	0.17	0
L:R106I	1.026	1.01	1.01	1.04	1.05	0.98		0

H:Q23K	1.183	1.35	0.97	1.09	1.26	1.07		1
H:R28T	0.376	ND	ND	0.81	ND	ND		-1
H:S30T	0.704	1.95	2.13	0.95	0.94	1.02		-1
H:N31S	0.365	ND	ND	0.84	ND	ND		-1
H:F32Y	0.925	0.96	1.01	0.92	0.92	1.02		1
H:V33A	1.250	0.81	0.55	0.99	0.85	0.82		0
H:I34M	1.089	1.31	1.16	1.10	1.26	1.10		-4
H:F45L	1.093	1.11	1.03	1.09	1.11	1.04		-2
H:P52aA	0.830	0.83	1.14	0.89	0.80	1.04	0.65	-5
H:Y53G	0.112	ND	ND	0.60	ND	ND	0.03	0
H:K57T	0.934	1.15	1.24	1.03	1.12	1.11		1
H:E58K	1.301	1.15	0.94	1.22	1.14	1.19		1
H:F59Y	0.847	0.74	0.98	0.87	0.76	1.00		6
H:A61Q	1.085	1.19	1.01	1.08	1.16	1.02		1
H:D65G	1.248	1.28	0.94	1.16	1.24	1.06		-1
H:F69I	1.095	1.08	1.07	1.10	1.09	1.03		-3
H:A71R	1.234	1.19	0.90	1.20	1.20	0.98		-3
H:N76S	1.278	1.17	0.80	1.15	1.19	0.91		0
H:R82aS	0.911	0.89	1.08	0.95	0.88	1.02		-1
H:A85E	0.913	0.88	1.07	0.94	0.89	1.08		4
H:S99C	0.790	0.68	1.11	0.80	0.62	1.03	0.9	-2
H:W100G	0.701	0.68	1.33	0.79	0.63	1.10	0.0	0
H:D100aG	0.664	1.02	2.20	0.87	0.71	1.29	0.4	1
H:I110T	1.083	1.00	1.03	1.09	1.02	1.08		-2

As may be seen above, differences from mature b12 observed with the HEK-expressed, purified germline reversion panel are significantly more modest than those previously observed with bacterially expressed F_{AB} constructs. Results for other CDR residues are, similarly, more drastic than seen in the HEK-expressed germline reversion mutants; given the discrepancies above, it is unclear whether this

results from experimental artifacts of the Protein G purification process, or whether the CDRs indeed display a high density of hot spot residues. In either event, the comparatively minor effects seen in the current panel stand in contrast to the dramatic mutational effects previously seen with b12.

SF162 vs. QH0692 Binding and Projected Cumulative Mutation Effects

The acquisition of binding kinetics and the modest per-residue changes in binding response facilitate comparison between the relative binding modes of b12 for SF162 vs. QH0692, and extrapolation of the projected effects of cumulative mutations to b12 along its maturation pathway. A point of particular interest is the fact that QH0692 has both slower association and dissociation rates than SF162 (2.8E5 $M^{-1}s^{-1}$ vs 9.6E5 $M^{-1}s^{-1}$ and 1.3E-3s^{-1} vs. 1.1E-2s^{-1} respectively; the typical interpretation of a decreased on-rate in protein:protein interactions is that the interface involved requires more extensive pre-ordering to bind, while the corresponding interpretation for a decreased off-rate is that the interaction is strengthened. This suggests that QH0692 should be making additional contact with b12, but that this additional contact requires more extensive conformational sampling on the part of gp120. A possible piece of supporting evidence for this is the decreased tendency of germline reversion mutants to lose binding to QH0692 (Fig. 5.6); even Y53G, by far the worst binder to SF162, generates fully 60% of the wild-type binding response. One explanation for this phenomenon would be that the additional contacts formed by QH0692 benefit from the decreased side-chain interaction from these mutants by limiting the conformational rearrangements required of QH0692 to bind, ameliorating the loss of affinity seen in SF162.

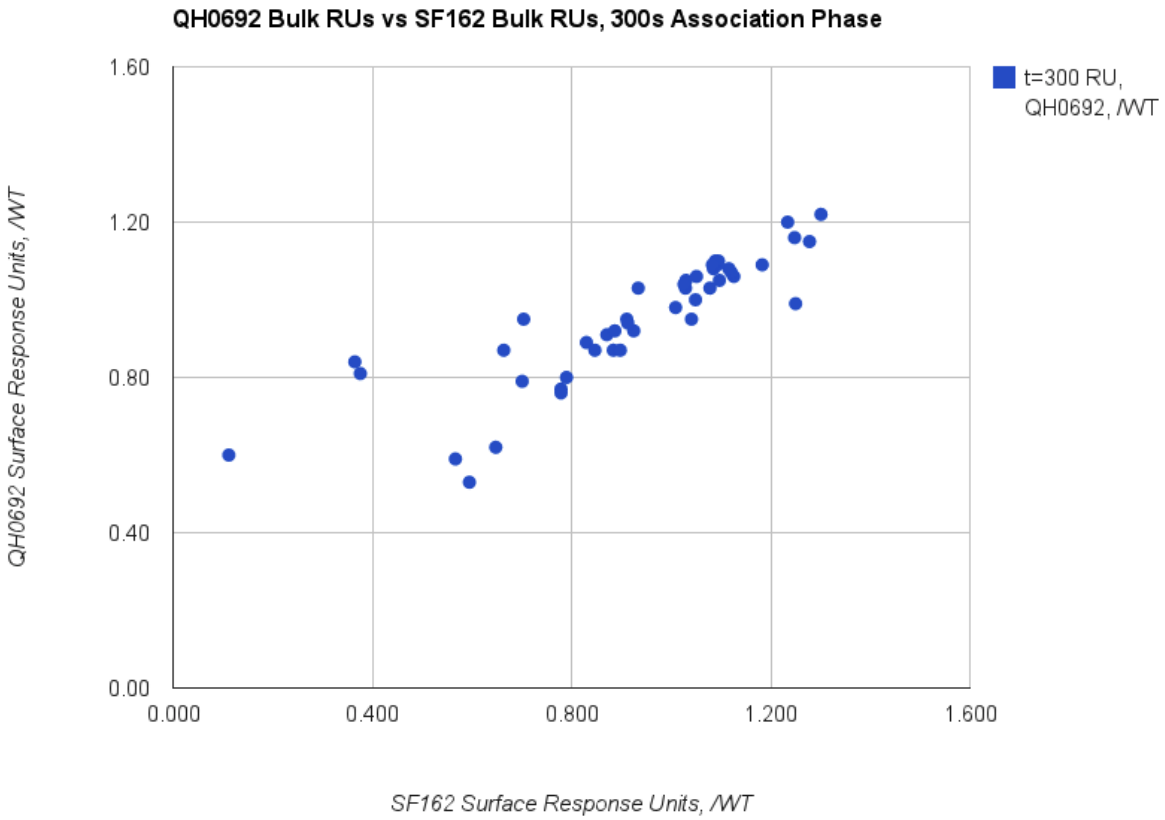


Figure 5.6: Effects on QH0692 vs. SF162 Surface Binding by Germline Reversion Mutations. While three mutations (H:R28T, H:N31S, and H:Y53G) reduce binding to SF162 by a factor of greater than two, none do so against QH0692.

Another very interesting question is: given that affinity maturation drives a low affinity, germline-like antibody towards a higher-affinity state, and that b12 appears to lack insurmountable barriers in affinity from any individual germline reversion mutation, what gains in biophysical parameters has maturation bestowed on b12? If we assume that the mutations to b12 are approximately additive, and that our measurement process is error-free, the product of all ratio effects on binding implies that a germline b12 scFv should:

- bind 5.6E2 times more weakly to SF162 as a heterogeneous surface
- bind 7.6 times more weakly to QH0692 as a heterogeneous surface
- associate 3.6 times more slowly with SF162, and 26 times more slowly with QH0692
- dissociate 26 times faster from SF162, and 8 times faster from QH0692

- bind 93 times more weakly to SF162 and 211 times more weakly to QH0692 using the calculated affinity for the high-affinity, fast-association component from the observed ka/kd ratio.

The highly disparate affinities for the heterogeneous surface and calculated dissociation constant for both gp120 species suggest, of course, that the situation is more complicated.

Discussion

Germline Reversions to b12 Are Dissimilar from Prior Mutagenesis Results

For both gp120 sequences assayed, our results indicate modest effects for mutations between the germline and mature b12 sequences. This result, while compatible with the somatic hypermutation process as typically understood, contradicts certain assumptions regarding anti-HIV BNAbs, and in particular regarding b12. First, the failure of projected b12 germline sequences to detectably bind gp120 has previously been interpreted as a sign that non-HIV stimulation of b12 progenitor sequences must have preceded subsequent gp120-driven somatic hypermutation; however, the modest effects of mutations towards the QH0692 sequence in particular suggest that, on the contrary, the effects on gp120 binding may be largely additive and cumulative, and that selection of the proper germline-stimulating gp120 sequence may elicit a b12-like development pathway. While alanine scanning results have indeed identified key interactions between b12 and gp120, these interactions are apparently less disrupted by mutation to the germline residue identities than to the alanines used in prior studies. Similarly, mutations previously discovered to have large impacts on binding have failed to differ between the recently discovered b12 germline precursor sequence and the mature b12 antibody; while these residues' identities may indeed be necessary for binding, antibody affinity maturation is unnecessary for their acquisition.

Second, the presence of a heterogeneous gp120 surface with an apparent high-affinity component - which may be partially selected for by the use of short-time-scale experiments - suggests a mechanism by which the generally inferior binding properties of a minimally mutated b12 germline sequence may be sufficient for proliferative stimulation of its parent B cell. While the determination of the precise variations in our gp120 surface composition responsible for the presence of a high-affinity

component is outside the scope of the current study, any such component reduces the threshold of affinity gain necessary for b12 germline stimulation.

Ramifications for CD4-Binding Site Antibody Elicitation

Aside from the important considerations drawn above from the lack of key hotspot mutations along b12's maturation pathway, a key finding of the present work is the cumulative importance of mutations without direct antigen contact in achieving high-affinity antibody binding. Previous work had identified heavy-chain-mediated residues with direct gp120 contact as potential mutations of interest; the finding that these contacts are generally preserved by the germline progenitor sequence implies that any initial, priming immunogen must present a surface with high complementarity to desirable CD4bs-contacting residues, and attempt to steer the rest of the antibody sequence towards desirable supporting mutations, possibly through additional contact surfaces outside the CD4bs.

This constraint is perhaps less critical for the much-studied VRC01-related antibody family, which possess germline binding to gp120 and have been demonstrably stimulated in a proliferative manner by gp120-derived constructs[66]; however, efforts to design away secondary contacts formed between immunogens and non-CDR antibody surfaces may be counterproductive as immunological priming reagents. One particularly interesting area of follow-on work is the attempt to use our new knowledge of the cumulative impact of b12 maturation mutations to design gp120-derived proteins with expected binding to germline or near-germline b12 sequences. In parallel, we will endeavor to find minimally mutated b12 germline sequences combining the most promising mutants from the panel assayed here, and assay their ability to bind gp120.

Materials and Methods

Preparation of b12 Point Mutant Reversions

Genes corresponding to the variable fragments of the wild-type, germline progenitor and heavy/light chain chimeric b12 sequences were codon-optimized for high synthetic output in human cells, ordered as synthetic genes from GenScript, and subcloned into the RKS17 lentiviral expression vector[67] or RKS33 CMV-early mammalian transfection vector. Point mutants were obtained for 44 of 45 reversions from the mature towards the germline sequence by primer mutagenesis (Table 6.2), with primer melting

temperatures determined by the formula T_m = 81.5 + 41*(G+C)/L - 675/L - 100*mismatches/L. One point mutant located in an A/T-rich region of the wild-type construct was obtained by Seamless assembly of two dsDNA fragments ordered commercially from IDT. Sequences were verified by BigDye gel electrophoresis, and stocks of plasmid for viral preparation or transfection generated via BenchPro 2100 automated Maxi prep.

b12_mature_LightLinkerHeavySTREP_ScFv:

CTCGAGATGCCTCTGGGGCTGCTGTGGCTGGGGCTGGCTCTGCTGGGGGCTCTGCACGCTCAGGC
TCAGGATGAAATCGTGCTGACTCAGTCCCCAGGAACTCTGTCTCTGAGTCCCGGCGAGAGAGCCAC
CTTCAGTTGCAGGAGCTCCCACTCTATCCGCAGTCGGAGAGTGGCCTGGTATCAGCACAAGCCTGG
CCAGGCTCCTAGGCTGGTCATCCACGGCGTGAGCAACAGAGCCTCAGGAATTAGCGACAGGTTCTC
AGGGAGCGGCAGCGGCACAGATTTTACTCTGACCATTACACGGGTGGAGCCTGAAGACTTCGCACT
GTACTATTGCCAGGTGTACGGCGCCTCTAGTTATACCTTTGGGCAGGGAACAAAGCTGGAGCGAAA
GCTGGTGCCCAGAGGCAGCGGAGGAGGAGGACTGGTCCCTCGCGGCAGTCAGGTGCAGCTGGTC
CAGTCAGGCGCTGAAGTGAAGAAACCAGGGGCAAGTGTGAAAGTCTCATGTCAGGCCAGCGGCTAC
CGATTCAGCAACTTTGTGATCCACTGGGTCCGACAGGCTCCAGGACAGCGGTTCGAGTGGATGGGC
TGGATTAATCCCTATAACGGGAATAAGGAGTTCAGCGCTAAATTTCAGGACCGCGTGACTTTTACCG
CAGATACATCTGCCAACACTGCTTACATGGAGCTGAGGTCCCTGCGCTCTGCAGATACCGCCGTGTA
CTATTGTGCAAGAGTCGGCCCTTATTCCTGGGACGATTCTCCACAGGACAATTACTATATGGATGTGT
GGGGCAAGGGGACCACAGTGATCGTCTCAAGCTCCGCTTGGTCCCATCCCCAGTTTGAAAAATGAG
GATCC

Mutation	Forward Primer
F43L	CGAGAGAGCCACCCTCAGTTGCAGGAGC
S47A	CCTTCAGTTGCAGGGCCTCCCACTCTATCCG
H49Q	GCAGGAGCTCCCAGTCTATCCGCAGTC
I51V	GCTCCCACTCTGTCCGCAGTCGGAG
R52S	CCCACTCTATCAGCAGTCGGAGAGTGGC
R54S	CCACTCTATCCGCAGTTCGAGAGTGGCCTGG
R55Y	CTCTATCCGCAGTCGGTACGTGGCCTGGTATCAGC
V56L	GCAGTCGGAGACTGGCCTGGTATCAGC
V70L	GCTCCTAGGCTGCTCATCCACGGCG
H72Y	CCTAGGCTGGTCATCTACGGCGTGAGC
V74A	GGTCATCCACGGCGCGAGCAACAGAG
N76S	CGGCGTGAGCAGCAGAGCCTCAGG
S79T	GTGAGCAACAGAGCCACAGGAATTAGCGACAG
S82P	GAGCCTCAGGAATTCCCGACAGGTTCTCAGGG
T99S	CTCTGACCATTTCACGGGTGGAGCCTGAAG
V101L	CTCTGACCATTACACGGCTGGAGCCTGAAG
L108V	GCCTGAAGACTTCGCAGTGTACTATTGCCAGG
V113Q	GCACTGTACTATTGCCAGCAGTACGGCGCCTC
A116S	GCCAGGTGTACGGCTCCTCTAGTTATACCTTTGG
R129I	GGAACAAAGCTGGAGATAAAGCTGGTGCCCAGAGG
Q169K	GTGAAAGTCTCATGTAAGGCCAGCGGCTACCGA

R174T	CAGGCCAGCGGCTACACATTCAGCAACTTTGTGATC
S176T	GCGGCTACCGATTCACCAACTTTGTGATCCAC
N177S	GCTACCGATTCAGCAGCTTTGTGATCCACTGG
F178Y	CCGATTCAGCAACTATGTGATCCACTGGGTCC
V179A	CCGATTCAGCAACTTTGCCATCCACTGGGTCCG
I180M	CAGCAACTTTGTGATGCACTGGGTCCG
F191L	CCAGGACAGCGGCTCGAGTGGATGG
P199A	GGGCTGGATTAATGCCTATAACGGGAATAAGGAG
Y200G	GGGCTGGATTAATCCCGGTAACGGGAATAAGGAGTTC
K204T	GGATTAATCCCTATAACGGGAATACGGAGTTCAGCGC
E205K	CCCTATAACGGGAATAAGAAGTTCAGCGCTAAATTTCAG
F206Y	CGGGAATAAGGAGTACAGCGCTAAATTTCAGGACC
A208Q	GGGAATAAGGAGTTCAGCCAGAAATTTCAGGACCGCGTGAC
D212G	CGCTAAATTTCAGGGCCGCGTGACTTTTACCG
F216I	GGACCGCGTGACTATTACCGCAGATACATCTG
A218R	CCGCGTGACTTTTACCAGAGATACATCTGCCAACACTG
N223S	GCAGATACATCTGCCAGCACTGCTTACATGGAG
R230S	CATGGAGCTGAGCTCCCTGCGCTCTG
A235E	GTCCCTGCGCTCTGAAGATACCGCCG
S249C	GTCGGCCCTTATTGCTGGGACGATTCTCC
W250G	GGCCCTTATTCCGGGGACGATTCTCCAC
D251G	GGCCCTTATTCCTGGGGCGATTCTCCAC
I270T	GGGACCACAGTGACCGTCTCAAGCTCC

Recombinant protein was expressed in HEK293F cells by either transient transfection or stable transduction by pseudolentivirual particles. Transient transfection was performed upon batches of 6E7 HEK293F cells by resuspension in fresh FreeStyle media to 2E6 cells/mL and the addition of 3 ug/mL RKS33 cDNA and 9 ug/mL linear polyethylimine; cells were fed an additional 30 mL of fresh media 24 hours post-transfection. Media was separated from cells 4 days post-transfection by centrifugation, and passed through high-capacity 0.22 um syringe filters.

Non-replicatively competent lentiviral particles were prepared by transient co-transfection of HEK293T cells with mutant scFv vector and two additional helper plasmids, with pseudoviral particles collected from the cell culture supernatant two days post-transfection and concentrated by overnight centrifugation at 8000*g*. Pseudoviral particles were resuspended in 400 uL () following aspiration of the cell culture media, and frozen in 100 uL aliquots of ~125x pseudovirus; transduction of HEK293F cells with frozen pseudovirus was performed using 1E7 cells in 10 mL of fresh FreeStyle media, with 20 mL fresh media added 18-24 hours post-transduction; 48 hours post-transduction, 15 mL of cultured cells

were pelleted by centrifugation (5 minutes, 1000g), washed once with pre-warmed media, and resuspended in 30 mL media for subsequent passaging and protein production.

Surface Plasmon Resonance Analysis

SPR experiments were conducted on a Biacore T-100. Protein samples were diluted by a factor of 10 or greater in HBS-EP buffer from sterile packaging immediately prior to use, and passed through a 0.22-μm-pore filter. Surfaces for b12 binding were created by amine coupling to lysines on gp120s purchased from ImmuneTech; samples were run in parallel against both gp120 sequences and a blank lane for proper double-referenced data. All scFvs were run in duplicate, with control injections of mature b12 scFv interspersed at least once per nine mutants.

Acknowledgements

The sequence used for b12 was designed in conjunction with Mesfin Gewe, who was also immeasurably helpful with getting the initial protein expression going. Colin Correnti was instrumental in trouble-shooting transient HEK expression issues, and Tim Vanden Bos performed the lentiviral particle generation steps in collaboration with Colin. SPR analysis was performed by Della Friend.

This work is intended to become part of a larger paper, in conjunction with crystal structures of a chimeric germline heavy/mature light chain b12 scFv and a dimer of germline light chains; the aggregate data should be sufficient to draw further inferences about anti-HIV antibody maturation in general, and about b12 in particular.

References:

[57] Hoot S et al, *PLoS Path.* 2013. 9(1):e1003106.
[58] Clackson T et al, *J Mol Biol* 1998. 277(5):1111-28.
[59] McHugh L et al, *J Biol Chem* 2002. 277(37):34383-90.
[60] Burton DR et al, *PNAS* 1991. 88(22):9705-6.
[61] Saphire EO et al, *Science* 2001. 293(5532):1155-9.
[62] Zhou T et al, *Nature* 2007. 445(7129):732-737.
[63] Zwick MB et al, *J Virol* 2003. 77(10):5863-5876.
[64] Azoiței ML et al, *Science* 2011. 334(6054):373-6.
[65] Barbas CF et al, *PNAS* 1994. 91(9):3809-13.
[66] Jardine J et al, *Science* 2013. 340(6133):711-716.
[67] Bandaranayake et al, *Nucleic Acids Res.* 2011. 39(21):e143.

Chapter 6: Conclusions and Future Directions

What Crystallization and Computational Design Successes and Failures Tell Us About Antibody:Antigen Interactions

The work presented over the previous four chapters focused largely on methods - computational protein design and protein crystallography - which present as their final outputs single structures which are in turn used to explain experimental observations from a solution environment closer to *in vivo* behavior than the artificial *in silico* or crystallographic environments necessary for model generation. While this property is often leveled against both methods as a criticism of their applicability, particularly in informal settings outside peer-reviewed publication, the tractability of particular proteins for computational structure prediction and crystallography can inform us of important properties of underlying biochemical systems. This is particularly true for antibody:antigen interactions, where the protein:protein interaction observed is the key determinant of clinically relevant biological function.

In all four of the research projects presented as components of this dissertation, the conformational flexibility of an antibody or antigen has limited or otherwise affected the interpretations attributable to a model often described in terms of fixed-geometry crystal structures:

- The loop insertion screening necessary to develop computationally designed V1/V2 scaffolds in Chapter 2 into crystallizable proteins reinforces the general flexibility of the V1/V2 loop even as the two successful solutions illustrate a stable V1/V2 conformation mediating PG9's interactions with gp120.
- In Chapter 3, EPIMAP fails to identify immunogenic stretches of gp120 where and only where the underlying conformational ensemble generated for the project did not sample a diverse structural ensemble, implying that both the identified and the unidentified antibody-accessible regions depend upon conformational flexibility for effective display to B cells.
- The failure of 2BODX scaffolds in Chapter 4 to entirely recapitulate their highly gp120-like bound epitope conformation despite high-affinity binding identifies a key weakness of affinity as a selection surrogate for structure, even as it reassures us that an initial template is sufficient to design a protein sampling a template-like structure sufficiently to recapitulate it when in complex with a desired binding partner.

- Mutational analysis of b12 reversions to germline demonstrated, in Chapter 5, a prior failure to appreciate the cumulative impact of minor mutations in the affinity maturation pathway of an anti-HIV broadly neutralizing antibody, supporting the idea that germline progenitor interactions are fundamentally similar to those of the mature antibody while drawing attention to the importance of the antibody light chain in bolstering heavy-chain-mediated antibody:antigen contacts.

From the results above, we can refine our understanding of the conventional model of antibody:antigen interactions in a more general sense. While a single, essentially fixed antibody:antigen complex conformation appears to mediate antibody interactions, neither the antibody nor the antigen appears to constantly occupy the bound conformation even in the case of high-affinity interactions which have been heavily optimized by somatic hypermutation. Thus, only select high-affinity gp120:PG9 interactions are sufficient for crystallization, and a collection of mutations with individually minor effects on structure and affinity mediate b12's development of high-affinity binding.

Chapters 2-4, in particular, represent an additional opportunity for future projects in that all use structure-based methods to assist in narrowing an intractably large problem space created by a protein with a problematic degree of flexibility. Given the increasing degree of biochemical literature available on the many interesting behaviors of intrinsically disordered proteins, and the increasing automation of more canonical problem types by the expansion of structural genomics consortia and publicly available modeling software, leveraging the limited structural resources available to bring edge-case problems within range of solution will continue to grow in importance as a subsection of the larger structural biochemistry field.

An example of such a project, in the form of a paper recently published in *PLoS Pathogens*, is attached to this dissertation as Appendix I. In collaboration with the Smith group at Seattle Biomedical Research Institute, and working in particular with Sowmya Sampath, I used Rosetta-based design methods to design minimally invasive glycan additions onto a malarial red blood cell invasion protein, despite having only a relatively low-resolution crystal structure of a homologous protein from a different species of malaria available for a design template; antibody responses elicited by the designed protein had improved neutralizing potency, as assayed by the inhibition of the malarial protein's binding to a

soluble form of its usual target receptor. This study represents an important proof of principle for vaccine design; while glycosylation has long been implicated in antibody targeting (as per chapters 2 and 3), antibody response focusing by an altered antigen had never been demonstrated. Attempts to improve HIV *Env* immunogenicity are at once both potentially high-reward, due to the disease's critical importance, and overwhelmingly likely to fail, given the virus's extensive mechanisms for immune evasion. Using other pathogens as easier test cases allows the testing of basic vaccine research concepts without the heavily entangled mess of potential failure modes intrinsic to HIV.

In summary, a common element among current applications of macromolecular structural methods towards HIV and other high-profile pathogens is their use in creating interpretable models from highly complex data; both computational modeling and crystallography allow the distillation of key determinants of experimental and, possibly, clinical success for candidate vaccines, including initial design, pre-clinical evaluation for suitability, and post-clinical evaluation for efficacy. The application of as broad a spectrum of complementary methods as possible to a problem increases the probability of finding a successful solution, especially in the context of HIV, where more conventional attempts at vaccine design have thus far failed.

Appendix I: Glycan masking of *Plasmodium vivax* Duffy binding protein for probing protein binding function and vaccine development

As submitted for publication in PLoS Pathogens (2013, 9(6):e1003420)

Sowmya Sampath[a*], Chris Carrico[b,c*], Joel Janes[a,], Sairam Gurumoorthy[a], Claire Gibson[a], Martin Melcher[a], Chetan E. Chitnis[d], Ruobing Wang[a], William R. Schief [b,e,f,#], Joseph D. Smith[a,g,#]

a. Seattle Biomedical Research Institute, 307 Westlake Ave N, Suite 500, Seattle, Washington, USA

b. Department of Biochemistry, University of Washington, Seattle, Washington, USA

c. Basic Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA

d. International Centre for Genetic Engineering and Biotechnology (ICGEB), New Delhi, India

e. Department of Immunology and Microbial Science and IAVI Neutralizing Antibody Center, The Scripps Research Institute, La Jolla, California, USA

f. Center for HIV/AIDS Vaccine Immunology and Immunogen Discovery, The Scripps Research Institute, La Jolla, California, USA

g. Department of Pathobiology, University of Washington, Seattle, Washington, USA

***** These authors contribute equally.

****** Corresponding authors. Joseph D. Smith, joseph.smith@sbri.org; William R. Schief, schief@scripps.edu.

Running Title: Glycan masking of *P. vivax* Duffy binding protein

KEY WORDS: *Plasmodium vivax*, malaria, Duffy binding protein, vaccines

1

Abstract

Glycan masking is an emerging vaccine design strategy to focus antibody responses to specific epitopes, but it has mostly been evaluated on the already heavily glycosylated HIV gp120 envelope glycoprotein. Here this approach was used to investigate the binding interaction of *Plasmodium vivax* Duffy Binding Protein (PvDBP) and the Duffy Antigen Receptor for Chemokines (DARC) and to evaluate if glycan-masked PvDBPII immunogens would focus the antibody response on key interaction surfaces. Four variants of PVDBPII were generated and probed for function and immunogenicity. Whereas two PvDBPII glycosylation variants with increased glycan surface coverage distant from predicted interaction sites had equivalent binding activity to wild-type protein, one of them elicited slightly better DARC-binding-inhibitory activity than wild-type immunogen. Conversely, the addition of an N-glycosylation site adjacent to a predicted PvDBP interaction site both abolished its interaction with DARC and resulted in weaker inhibitory antibody responses. PvDBP is composed of three subdomains and is thought to function as a dimer; a meta-analysis of published PvDBP mutants and the new DBPII glycosylation variants indicates that critical DARC binding residues are concentrated at the dimer interface and along a relatively flat surface spanning portions of two subdomains. Our findings suggest that DARC-binding-inhibitory antibody epitope(s) lie close to the predicted DARC interaction site, and that addition of N-glycan sites distant from this site may augment inhibitory antibodies. Thus, glycan resurfacing is an attractive and feasible tool to investigate protein structure-function, and glycan-masked PvDBPII immunogens might contribute to *P. vivax* vaccine development.

Author Summary

An important goal of many vaccine efforts is to inhibit pathogen invasion of host cells, but few approaches exist to target vaccine antibodies on invasion blocking epitopes. Glycan masking is a vaccine design strategy to hide protein surfaces with carbohydrates and focus antibodies on exposed surfaces. This approach has mostly been evaluated on the heavily glycosylated HIV envelope glycoprotein, but it has never been tested on eukaryotic pathogens, such as *Plasmodium*, which have limited N-glycosylation machinery and therefore may provide a better platform to explore this strategy. Here, we used glycan masking to investigate the binding interaction between *Plasmodium vivax* Duffy binding protein (PvDBP) and the Duffy Antigen Receptor for Chemokines (DARC). This study showed that addition of an N-glycan site in a predicted host interaction surface abolished binding and potentially covered up an inhibitory antibody epitope. In contrast, addition of multiple N-glycan sites distant from predicted interaction surfaces did not inhibit binding but did slightly enhance elicitation of inhibitory antibodies. This analysis shows that glycan resurfacing offers an integrated approach to characterize protein function and immunogenicity and that glycan resurfacing of PvDBPII immunogens may have utility in *P. vivax*-malaria vaccine development.

Introduction

Plasmodium vivax invasion of human reticulocytes is strongly dependent on an interaction between the *P. vivax* Duffy Binding Protein (PvDBP) and the Duffy Antigen Receptor for Chemokines (DARC) on the reticulocyte surface [1]. DARC-negative individuals are highly resistant to *P. vivax* infection [2] and the DARC-null phenotype has independently arisen in different human populations [3,4]. Although an alternative pathway of *P. vivax* invasion has recently been described [5,6], DARC-null carriers have reduced susceptibility to *P. vivax* infection [4,7] and the FyA DARC allele shows reduced binding to PvDBP and is more susceptible to antibody blocking [8]. Thus, the PvDBP-DARC interaction has a critical role in *P. vivax* infection making it an attractive vaccine target.

The molecular mechanisms of PvDBP-DARC binding and immune evasion are only partially understood. PvDBP is a member of the Erythrocyte Binding-Like (EBL) protein superfamily [9–11]. The extracellular region of PvDBP has been divided into six regions [10], of which DARC binding has been localized to region II (PvDBPII) [12]. The structure of PvDBPII [13] and a related Duffy binding protein of the simian malaria parasite *P. knowlesi* (Pkα-DBL) [14] has been solved and is composed of three subdomains – subdomain 1, 2 and 3. PvDBPII binds to the N-terminal 65 residues of DARC with a sulfated tyrosine on DARC at position 41 having a critical role in binding [15,16].

Although the PvDBP structure has been solved, the precise extent of the DARC binding "footprint" remains unclear [17,18], and there is limited understanding of the epitopes for DARC-inhibitory antibodies on PvDBP. Two different PvDBP-DARC binding models have been proposed (Figure 1). The "just in time" model hypothesizes that PvDBP engages DARC in a monomer-monomer interaction and that binding occurs so rapidly that the binding site is not under strong antibody attack [14]. In this model, the putative sulfotryosine binding pocket is located at a relatively flat surface in subdomain 2, on the opposite surface from a cluster of polymorphic residues [1]. It has also been proposed that adjacent residues from subdomain 1 form the sulfotyrosine-binding pocket, in an analogous manner to how sulphated tyrosines facilitate the gp120-CCR5 interaction during HIV invasion [19]. For convenience, we will refer to the flat surface on subdomain 2 as the sulfotryosine binding pocket (STBP) when referring to this location (light blue residues in Figure 1), although this remains to be established. The second model, termed "receptor-mediated ligand dimerization", hypothesizes that DARC binding drives dimerization between PvDBP

monomers and that the sulfotyrosine-binding pocket and DARC binding pocket are formed at the dimer interface, at distinct residues from the “just in time” model [13].

These two binding models are not necessarily exclusive but lead to different predictions about how antibodies may inhibit *P. vivax* invasion of reticulocytes. High titer and broadly neutralizing antibodies to PvDBP are only found after repeated infection [20–23], indicating that PvDBP is not highly immunogenic during infection. Naturally acquired inhibitory antibodies map to the dimer interface [13,20], suggesting that one mechanism of antibody action may be interference with receptor dimerization. Other inhibitory epitopes have been mapped to the highly conserved subdomain 3 [24] and a polymorphic residue located near the putative STBP in subdomain 2 [25], indicating that some inhibitory epitopes may be selected for immune evasion. *P. vivax* dependence on DARC for invasion and its restrictive preference for reticulocytes supports PvDBP vaccine development. However, the rapid kinetics of parasite invasion (less than one minute) [26,27] combined with PvDBP polymorphism [28,29] pose challenges for vaccine development that may be addressable by the rational redesign of PvDBPII for improved immunogenic properties.

Structural vaccinology is an emerging strategy that uses three dimensional structure to guide immunogen design to improve functional antibody responses. Viral enveloped pathogens, like HIV, use a number of mechanisms to evade immune surveillance, one of them being glycan shielding of epitopes [30]. Glycan masking, or surface glycoengineering, has been applied to HIV vaccine development as a means to shield regions of the gp120 surface from antibody recognition, but this has not resulted in elicitation of more potent or more broadly-cross-reactive neutralizing antibody responses [31–35]. However, gp120 is one of the most heavily glycosylated proteins in nature, so it may not offer the optimum protein scaffold to evaluate this strategy. Conversely, many deep branching eukaryotic pathogens, such as *Plasmodium* and *Giardia*, have limited N-glycosylation machinery [36], and consequently are not as highly evolved for glycan shielding. Thus, malaria proteins may be more amenable to glycan masking. Here, glycan masking was employed as a strategy both to investigate the PvDBP-DARC binding model and to evaluate whether it increased DARC-binding-inhibitory antibodies.

Results

Design of surface re-engineered PvDBPII recombinant proteins containing additional N-linked glycan sites

The monomeric crystal structure of the *P. knowlesi* α DBL domain [14] was used to guide glycan masking of PvDBPII recombinant proteins. Defined secondary structure regions in PvDBPII are largely helical [13] – subdomain 1 contains a β-hairpin, subdomain 2 is a four-helix bundle, and subdomain 3 contains a double helical bundle (Figure 1). N-glycosylation sites occur most frequently in regions of change in secondary structure, and least frequently on helices than any other secondary structural element [37]. The PvDBPII dimer interface is formed between contacts across helical domains in the two monomers [13], and consequently N-glycan acceptor sites were not added to this location. Instead, glycosylation sites were mostly added on surface exposed loops and the ends of helices that were not expected to contribute to invasion blocking epitopes but were expected to be accessible to antibodies.

The PvDBP sequence from the *P. vivax* Salvador 1 strain (Sal1) contains three potential N-glycosylation sites at positions 255, 351 and 420 (Figure 2). As the three native positions were located at the surface of subdomain 2 or subdomain 3 of PvDBPII distinct from the putative DARC interaction surface or dimerization surfaces (Figure 1), all three were retained in all of the hyperglycosylated variant sequences (Table 1). Three DBPII glycosylation variants were designed containing a total of four, six and 11 N-linked glycans, called “STBP glycan”, “P1” and “Max” respectively (Table 1, Figure S1). To evaluate the PvDBP-DARC interaction model [14], the “STBP glycan” was generated by adding an N-glycan site at position 374, where it would potentially cover the predicted sulfotyrosine binding pocket in subdomain 2 (model 1, Figure 1) and inhibit binding to reticulocytes. The remaining two glycan variants were designed to shield polymorphic residues in PvDBP but retain exposure of proposed interaction surfaces for DARC binding and/or PvDBP dimerization (Figures 1 and 2). There are at least 36 polymorphic positions in PvDBPII [28,29], but many sequence variants are extremely rare (Figure S2). The majority of polymorphic residues are dimorphic and localized to subdomain 2 (Figure 2). Critical DARC binding residues and several linear antibody inhibitory epitopes have been mapped to the N-terminus of subdomain 1, helices 2-5, and the connecting loop between subdomains 2 and 3 (Figure 2) [17–20]. P1, the second DBPII glycosylation variant was generated by adding three extra N-linked glycosylation sites at positions 264,

462, 486. The addition of an N-glycan site at 264, along with the native N- glycan sites at position 255 and 351 was designed to shield polymorphic residues that are present in subdomain 2 (Figures 1 and 2). N-glycan sites at position 462 and 486 along with the native glycosylation site at position 420 were designed to cover a portion of subdomain 3 (Figures 1 and 2). A third DBPII glycosylation variant, “Max”, contained the six N-glycan sites on P1, plus five additional N-glycan sites at 232, 341, 412, 467 and 495, thus bringing the consensus N-linked glycan sites in Max to 11. Addition of N-glycan sites at 412, 462, 467, 486, and 495, along with the native site at 420, were designed to cover all the surface accessible loops on subdomain 3. Further addition of N-glycan sites at positions 264 and 341, along with the native N-glycosylation sites at 255 and 351, were designed to cover all loops in subdomain 2 including a polymorphic loop between helices 1 and 2 and the highly polymorphic helix 5 (Figures 1 and 2). The Max glycosylation variant, with 11 N-glycans, should have a large surface area of PvDBPII shielded by glycosylation (Figure 1), but none of the N-glycosylation sites are at the dimer interface or predicted to interfere with DARC binding.

Expression and purification of surface re-engineered PvDBPII recombinant proteins containing additional N-linked glycan residues

To produce glycoengineered recombinant proteins, constructs were secreted from the mammalian cell line HEK293F. All of the recombinant proteins were purified to homogeneity, and no truncation products were seen (Figure 3A-B). Each N-glycan adds about 2 kDa to the protein. A ladder effect consistent with increasing N-glycosylation was observed for the recombinant proteins. The wild-type PvDBPII (3 N-glycan sites), STBP glycan and P1 ran at their expected molecular weight, whereas Max migrated slightly lower than the expected fully glycosylated molecular weight of 65 kDa (Figure 3A). Partial glycosylation (underutilization) of some of the 11 N-linked glycosylation sites could account for this. When the proteins were subjected to N-glycosidase treatment, all of DBPII glycosylation variants and the wild-type protein were reduced to a similar size (Figure 3C). These results suggest that many of the introduced sites were appropriately glycosylated in PvDBPII and its N-glycan variants.

Functional characterization of hyperglycosylated PvDBPII variants by standard in vitro binding assay

The functional effect of adding N-glycans on PvDBPII was assessed by using an established in vitro PvDBPII cytoadherence assay [12]. In this assay PvDBPII is expressed at the surface of COS-7 cells as a fusion protein with HSVgD1 to aid in surface expression [23]. A GFP reporter is added at the cytoplasmic tail to facilitate identification of transfected cells. Fluorescent cells covered with five or more RBCs, termed rosettes, are considered positive for binding [38]. As expected, nearly all of the cells transfected with wild-type PvDBPII bound to RBCs (Table 1 and Figure 4A). In contrast, addition of the single, extra N-glycan site at N374 nearly completely abolished binding of the STBP glycan mutant (4%) (Table 1 and Figure 4B), while both the P1 and Max variants exhibited binding similar to wild type PvDBP (Figure 4C and D). The reduced binding activity of the STBP glycan mutant was not due to defective surface expression, since it was surface expressed as well or better than the wild-type, P1, or Max glycosylation variants (Figure 4 and Figure S3 for flow cytometry gating strategy). Furthermore, although it was difficult to resolve if there was a difference between STBP glycan and wild-type DBPII in the COS-7 system, P1 and Max migrated slower than wild-type protein prior to N-glycosidase treatment indicating they were hyperglycosylated (Figure S4). Thus, consistent with prior data showing that mutation at I376 strongly reduces DARC binding [17], we find that introduction of the N374 glycan site (I374N/I376T) nearly abolishes DARC binding, either as a consequence of the mutation or addition of N-glycosylation. In addition, the P1 and Max mutants were functionally folded, despite additional surface glycoengineering.

A meta-analysis of previous PvDBP-DARC binding studies and the new glycan mutants was performed. The majority of previous site-directed mutagenesis was concentrated in subdomains 1 and 2 (Table S1). Fewer residues had been mutated in subdomain 3 and many of these were buried residues (Table S1). In this study, a total of two, three and seven surface residues were modified in subdomains 1, 2, and 3 with no effect on DARC binding. Overall, critical binding residues (30% or more loss of binding) are concentrated to the predicted binding sites in subdomains 1 and 2 (model 1) or at the dimer interface (model 2), suggesting both areas are critical for DARC binding. Conversely, no surface changes in the highly conserved subdomain 3 have affected DARC binding (Figure 1G, H and Table S1).

Immunogenic characterization of PvDBPII and its hyperglycosylated variants

To investigate immunogenicity, mice were immunized with wild-type and DBPII glycosylation variants. Since the yield of PvDBP recombinant proteins was low (between 0.1 to 0.5 mg/L) we employed a **DNA prime–protein boost** strategy (Figure 5A). In this approach, mice were immunized by DNA electroporation with plasmid DNA and boosted with the matching DBPII wild-type or DBPII glycosylated variant protein produced in HEK293F cells (Figure 5A). As a control, mice were immunized with wild-type protein alone that was produced in *Escherichia coli* (no N-glycosylation) or HEK293 cells (3 native N-glycosylation sites).

Antibody titers were compared using the *E. coli* recombinant protein that lacks N-glycosylation and therefore has no potential antibody epitopes hidden. One-way analysis of variance (ANOVA) of individual mouse ELISA titers showed no significant difference in antibody titer between any of the six immunization groups (Table S2) at the time of the final bleed (Figure 5B). This indicated that there was no loss of immunogenicity of glycoengineered recombinant proteins (groups 4-6), despite having significant surface glycosylation (Figure 1). Furthermore, mice that received four protein immunizations with 5 µg of protein per immunization (groups 1 and 2) had similar final antibody titers to mice that received four DNA immunizations followed by two protein boosts with 2.5 mg of protein per immunization (group 3). Thus, DNA electroporation**–protein** boost offers a strategy to circumvent issues that arise due to limited availability of recombinant PvDBPII protein. However, DNA electroporation alone was insufficient to elicit strong adhesion blocking antibody responses. The antibody titer increased by more than an order of magnitude after the two protein boosts (Figure 5B) and there was almost no inhibitory activity following the fourth DNA immunization (Figure 5C). Conversely, plasma obtained after the two protein boosts exhibited considerable inhibitory activity even at 1:1800 dilution in a standard COS-7 cell**-red blood cell (RBC)** binding inhibition assay [23] (Figure 5C).

Inhibition of PvDBPII binding to DARC on erythrocytes

Plasma from the different immunization groups were evaluated for inhibitory activity using the COS-7 cell**–RBC** binding assay. The inhibitory assay was carried out for plasma dilutions ranging from

1:200 to 1:10800 (Figure 6A and S4). A 1:200 dilution of pre-immune plasma showed no inhibition of RBC binding (rosette formation) and was used as control. A statistical comparison of the Least Squared Distance of data points shows that the non-linear model provides a good fit. Thus, the percent inhibition of rosette formation was plotted against log plasma dilution and analyzed by non-linear regression to determine the IC50. The IC50 values of the wild-type *E. coli* and HEK293 DBPII proteins were similar (Figure S5). To investigate the effect of glycan masking the IC50 values of wild-type PvDBPII and the three DBPII glycosylation variants in the DNA-protein prime boost regimen were compared. Individual values ranged from 1:2100 to 1:2800 and were not statistically different in the non-linear model (Figure 6A, Table S3). This result shows that DBPII glycosylation variants elicited similar levels of inhibitory antibodies to wild-type immunogen in the COS-7-RBC binding format, despite having significantly greater surface glycosylation.

Inhibition of PvDBPII binding to recombinant DARC

To further investigate whether any of the glycosylated immunogens provided a small improvement in inhibitory antibodies, we developed a novel flow cytometry based antibody-inhibition assay that is easier to quantify and more amenable to high-throughput analysis than the COS-7-RBC binding assay. In this assay, wild-type PvDBPII, minus the 3 native N-glycan sites, is expressed by yeast surface display [39] and probed with recombinant DARC-Fc. Binding of DARC-Fc is detected using an antibody to the Fc fusion domain. In control experiments, the wild-type YY-DARC-Fc recombinant protein bound to transfected yeast cells, while the YF-DARC-Fc variant lacking the critical sulfotyrosine site at position 41 [15] did not bind (Figure S6). The percent inhibition of wild-type DARC-Fc binding was then plotted against log plasma dilutions from the four immunization groups. In this assay, plasma from the STBP glycan variant had slightly weaker inhibitory activity than the wild-type PvDBP immunogen (Figure 6B), while P1 and Max elicited stronger inhibitory activity. The IC50 values for P1 and Max were 7-fold and 20-fold lower than for wild-type immunogen (Figure 6B, Table S3). This difference was statistically significant ($p=0.02$ for both P1 and Max, one tailed Student's *t* test). Moreover, the percent inhibition by P1 or Max plasma was significantly better than wild-type at multiple plasma dilutions (Figure 6B). This analysis suggests that an inhibitory epitope in PvDBPII was eliminated in the STBP glycan immunogen,

whereas increased levels of DARC-binding-inhibitory antibodies were elicited by P1 and Max DBPII glycosylation variants.

Effect of the FyA/B DARC allele on binding inhibition

A potential explanation for the discrepancy in antibody inhibition results between COS-7 and yeast display formats is that the yeast assay may be more sensitive to detect differences in antibody inhibitory activity. In the yeast assay, antibodies are challenged to inhibit an interaction of limited valency (the binding of DARC-Fc to yeast-displayed-PvDBPII is mono- or bi-valent) while in the COS-7 assay, antibodies are challenged to inhibit an interaction between two cells (RBCs displaying DARC and COS-7 cells displaying PvDBPII) that is likely to be multi-valent and potentially highly multi-valent (Figure 7). We initially conducted our assays using an FyB RBC donor (COS-7) or an FyB DARC recombinant protein (yeast) [15] . However, the FyA allele is more common in *P. vivax* endemic regions, and recent work suggests it has lower PvDBP binding activity and greater susceptibility to antibody blockade than FyB [8]. To investigate if the immune plasma would have greater activity on FyA donors in the COS-7 format, we screened five FyA/FyA donors and one FyB/FyB donor using the COS-7 cell**–RBC** binding assay. Surprisingly, only two of the five FyA donors exhibited reduced binding activity. This was mostly manifest as smaller RBC rosettes (Figure 8 inset). The COS-7 cell**–RBC** antibody inhibition assay was then repeated with RBCs from the weakest binding FyA donor and an FyB donor (Figure 8). Notably, a similar percentage of transfected COS-7 cells bound FyA and FyB RBCs. However, FyA RBCs formed smaller rosettes and were more susceptible to antibody inhibition from each of the four immunization groups than FyB cells (Figure 8). P1 plasma had measurably higher inhibitory activity than wild-type plasma at 1:200 and 1:600 dilutions against FyA and FyB RBCs, but this trend only achieved significance for one dilution value in the FyB experiments and was not apparent at higher dilutions. No equivalently significant trends were observable for STBP or Max plasma.

Discussion

Efforts to make a blood stage malaria vaccine have proven extremely challenging because of antigen polymorphism, redundant pathways of RBC invasion, and the rapid kinetics of parasite invasion [40]. Although *P. vivax* is highly restricted to reticulocytes and theoretically poses fewer challenges for blood stage vaccine development than *P. falciparum*, PvDBP is polymorphic [28,29] and vaccine-induced antibodies are only partially effective at inhibiting the PvDBP-DARC binding interaction [21,24,41–44]. Therefore, new immunogen design strategies and adjuvants [45] are required to enhance neutralizing antibody responses and confer cross-strain protection.

Glycan masking is an emerging vaccine design strategy to target antibodies to specific antibody epitopes. A glycan shield is employed by HIV and other enveloped viruses, such as influenza, to hide regions of viral surface proteins from neutralizing antibodies [46–48]. N-linked glycans are bulky, flexible structures, more than 20 times the size of an amino acid side-chain. Carbohydrates are poorly immunogenic, and the large carbohydrate unit effectively shields the underlying amino acids from antibody recognition [49–52]. Although some HIV-broadly-neutralizing antibodies isolated from natural infection utilize glycans as part or all of their epitope [39,53–56], the basis for glycan-targeting in these cases may be due to the extraordinarily high density of glycans on HIV Envelope (which may create “non-self” epitopes from “self” glycan molecules) and/or to the fact that HIV infection causes significant dis-regulation of the immune system or other factors. Indeed, glycan reactive monoclonal antibodies display extensive somatic hypermutation and/or unusually long complementarity determining regions [57,58], suggesting that their elicitation may require highly persistent infection. In the context of vaccine design, glycan masking has mostly been evaluated on HIV Envelope glycoproteins. Glycan masking has reduced monoclonal antibody recognition of targeted epitopes on HIV Envelope gp120/gp140, but it has not led to broader neutralizing antibody responses [32–34]. However, gp140 contains 30 potential N-glycosylation sites on average, and most of its surface is predicted to be covered by N-glycans [59]. In contrast, malaria encodes a minimal N-glycosylation machinery and malaria surface proteins may contain limited or no N-glycosylation. *P. falciparum* encodes an identifiable Alg7 glycosyltransferase and SST3 oligosaccharyltransferase, and it is predicted that only a short $GlcNAc_2$ group would be transferred to

proteins [36]. Thus, *Plasmodium* proteins may be better platforms to evaluate this vaccine design strategy because they are not already highly evolved to evade antibody responses via glycan masking.

Here, glycan masking was used both to probe PvDBP binding function and to test a vaccine strategy to focus antibodies on critical interaction surfaces. Although several different PvDBP-DARC binding mutagenesis studies have been done, there is still controversy surrounding the PvDBPII-DARC binding model [13,14]. The first mutagenesis studies were performed before the PvDBP structure was solved [17,18] and there has been limited structure-guided mutagenesis analysis. The present study shows that elements of both the “just in time” and “receptor-mediated dimerization” models are correct. We found that an N-glycan site added to a predicted sulfotryosine binding pocket in subdomain 2 abolished DARC binding, and that N-glycan sites added distant from this interaction site or the dimer interface had no effect on DARC binding. These findings confirm the importance of the subdomain 2 interaction site for DARC binding, as predicted in the “just in time” model, and extend our understanding of the DARC “binding footprint”. The meta-analysis also suggests that residues critical for DARC binding [13,17–19] are localized to the dimer interface (as predicted by the “receptor-mediated dimerization” model) and at a narrow patch on subdomain 2 and wrapping around to adjacent regions in subdomain 1 (consistent with the “just in time” model). It is possible that the DARC interface encompasses both regions or that the DARC binding site and dimer interface are separate and distinct, but both may contribute to efficient DARC binding. There are still many surface residues in close proximity to the predicted DARC interface in model 1 that need to be evaluated for binding, including the adjacent subdomain 3.

In mouse immunization studies, glycoengineered proteins were equally immunogenic to wild-type proteins by ELISA. This finding is encouraging because it suggests that substantial glycoengineering did not inadvertently blunt T helper cell priming [60,61] or the development of anti-PvDBP antibody responses. However, differentially glycosylated PvDBP variants failed to yield significant differences in antibody inhibitory activity in the COS-7-RBC binding assay. While the response of P1 plasma at 1:200 and 1:600 dilutions hints at improved response against the P1 variant, the difference in response was insufficient to allow proper determination of a significant difference in elicited IC50.

In contrast, consistent and significant enhancement of PvDBP:DARC binding inhibition was observed in the yeast display assay used in this study. Of particular interest is the observed increase in

inhibitory potency elicited by the P1 and Max constructs with increased glycosylation distal to predicted inhibitory antibody binding areas, as well as correspondingly decreased inhibition elicited by the STBP construct. The extra N-glycosylation site in the STBP construct covers a linear inhibitory antibody epitope from natural *P. vivax* infections [20], adding plausibility to the interpretation that inhibitory antibodies map near this region and can be targeted by vaccination. The yeast display assay provides greater ease and throughput and may be more sensitive to inhibitory antibody responses due to the lower PvDBP-DARC binding valency in the yeast format (Figure 7). This sensitivity to detect small differences in antibody activity may be important in engineering more potent PvDBP immunogens in the future. Indeed, small 2-fold differences in DARC surface levels [4] or FyA allelic polymorphism [8], are thought to affect PvDBP binding activity and the ability of anti-PvDBP antibodies to block parasite invasion (model in Figure 7). Thus, this interaction may be particularly vulnerable to antibody blockade.

If the P1 and Max variants successfully enhanced inhibitory binding titers as implied by the yeast display assay, critical questions for further study are whether the increased potency observed in the yeast assay is indeed clinically relevant, and whether the glycosylated constructs developed in this study are amenable to further engineering and concomitant improvement in elicited immune response. One possibility is that in designing the current constructs we may have inadvertently masked inhibitory epitopes that have recently been localized to subdomain 3 [24] and are only partially mapped [20] (Figure 2). In addition, it remains possible that the more pronounced activity of Max plasma in the yeast assay compared to the COS-7 platform may be a consequence of the fact that the PvDBP constructs in the COS-7 and yeast display assays were smaller than the PvDBP vaccine immunogen and slightly different from each other to facilitate protein expression (Figure S1). In particular, the N-terminus of the yeast display construct was 8 amino acids shorter than the COS-7 construct and 18 amino acids longer at the C-terminus. Thus, potentially inhibitory antibodies directed against the C-termini of PvDBP would be less likely to be detected in the COS-7 assay than the yeast assay. As more inhibitory epitopes are mapped in PvDBP, it may be possible to improve glycoengineered PvDBPII immunogens by repositioning N-glycosylation sites to shield off-target epitopes and focus antibodies on more strain-conserved inhibitory epitopes. Whereas this study used a DNA and protein prime-boost approach, glycoengineering could also

be applied with viral vectored vaccines, or with protein-only immunization if a strategy for improved PvDBP expression could be developed.

In conclusion, this analysis shows that PvDBPII recombinant proteins retain immunogenicity despite considerable glycoengineering and that glycan resurfacing offers an integrated approach to characterize protein function and immunogenicity. Glycan resurfacing of PvDBPII immunogens may have utility in *P. vivax*-malaria vaccine development.

343 **Materials and Methods**

344 **Cloning of PvDBPII in pTT3 vector**

345 PvDBPII and PvDBPII glycosylated variants were synthesized at Genscript with mammalian
346 codon optimization and cloned into a modified version of the pTT3 vector (kindly provided by Dr. Yves
347 Durocher, Biotechnology Research Institute, National Research Council Canada) using HindIII and
348 BamHI restriction sites. The PvDBPII is in frame with a signal sequence from murine IgG kappa chain at
349 the N-terminus and 6X His tag at the C-terminus [62]. Sequences of wild-type PvDBPII (Salvador 1 strain,
350 Sal1) and glycosylated variants are provided in Figure S1.

351 **Design of glycosylated variants**

352 N-linked glycosylation site selection for PvDBPII (Sal1 variant) was performed by fixed-backbone
353 redesign in Rosetta, using Rosetta-generated homology models derived from the crystal structure of the
354 *P. knowlesi* PkDBPII (2C6J) [14,63,64]. The PvDBPII sequence was checked for positions where one
355 mutation adds an N-glycan consensus sequence N-X-(S/T) where X is any amino acid except for proline
356 [65]. Putative glycan sites were screened for solvent accessibility of the glycan-bearing asparagine
357 residue; sites were mostly spaced no closer than 5 amino acids to an existing glycosylation site based on
358 prior work showing decreased efficiency for closely spaced sites [66]. Steric clashes due to N-glycans
359 being in proximity were addressed by using the Rosetta program. During site selection, consideration was
360 also given to secondary structure preferences for N-glycosylation sites [37], and to avoid the introduction
361 of buried unsatisfied polar groups at the S/T position.

362 **Expression and purification of PvDBPII and PvDBPII glycosylated variants**

363 HEK-293F cells were cultured in SFM4TRNSFX media (Hyclone). For expression and purification, 400
364 million cells were suspended in 20 ml of fresh SFM4TRNSFX media. Cells were transfected with 500 µg
365 of DNA and 1 mg of PEI MAX (Polysciences.com) using a previously described procedure [67]. This
366 suspension was shaken at 125 rpm and incubated at 37^{o}C and 5% CO_2 for 4 h, then made up to 400 ml
367 with Freestyle media (Invitrogen) and placed back in the shaker. Tissue culture supernatant was
368 harvested 5 days after transfection and spun at 1000 rpm for 20 min to remove cell debris. Following
369 centrifugation, the supernatant was balanced with 20 mM HEPES pH 7 and 200 mM sodium chloride.
370 Recombinant PvDBPII proteins were affinity purified using Nickel Sepharose 6 Fast Flow resin (GE

371 Healthcare), dialyzed into 1X PBS and stored. To assess purity, recombinant PvDBPII and glycosylated
372 variants were run on 4 to 20% SDS gel (Invitrogen) and stained with GelCode Blue solution (Invitrogen)
373 or transferred to PVDF membranes and probed with anti-His antibody or with rabbit anti-PvDBP plasma
374 for Western blots.

375 **Ethics statement**

376 This study was carried out in strict accordance with the recommendations in the Guide for the Care and
377 Use of Laboratory Animals of the National Institutes of Health. Rabbit immunizations were performed by
378 custom vendor at R&R Rabbitry. Mice immunizations were performed by Seattle Biomedical Research
379 Institute personnel. Animals were housed under controlled laboratory conditions. The protocols were
380 approved by the Institute Animal and Care Use Committee at Seattle Biomedical Research Institute
381 (Protocols JS-ABP and JS-04).

382 **Rabbit and mice immunizations**

383 For rabbit immunizations, four rabbits were immunized with *E. coli* generated refolded PvDBPII protein at
384 R&R Rabbitry in accordance with current guidelines on animal immunizations. Two animals received 50
385 µg of protein in Freund's complete adjuvant for the first injection followed by protein in incomplete
386 Freund's adjuvant for the subsequent 5 boosts. Two others received 50 µg of protein in TiterMax gold
387 adjuvant (TiterMax.com) for the immunization and the boosts. Boosts were carried out every 3 weeks. For
388 mice immunizations, 6 groups of 10 mice each were subjected to immunization according to animal
389 immunization guidelines. Group 1 and 2 mice received protein only for the immunization and three boosts.
390 Mice in groups 3 to 6 received DNA immunization followed by three DNA boosts and two protein boosts
391 at three week intervals. Protein immunizations were given as intramuscular injection into the left tibialis
392 muscle. DNA immunization was done by DNA electroporation in the anterior tibialis muscle using Trigrid
393 delivery system (ICHOR). Group 1 mice received 10 µg of *E. coli* generated, refolded PvDBPII. Group 2
394 mice received 5 µg of PvDBPII generated in HEK-293F cells. Groups 3-6 mice received 20 µg of DNA
395 and 2.5 µg of PvDBPII, PvDBPII STBP glycan, PvDBPII partial glycosylated variant (P1) and PvDBPII
396 maximum glycosylated variant (Max), respectively (Table S2). A combination of CpG Oligo dinucleotide
397 (ODN1826, Invitrogen) and alum was used as the adjuvant for protein immunizations [68].

398 **ELISA**

100 ng of recombinant PvDBPII produced in *E. coli*, refolded, and lacking in glycosylation was coated on 96 well plates (Nunc) by overnight incubation at 4^{o}C. Mice plasma dilutions ranging from 1:500 to 1:512,000 were added to the plate and probed with ELISA kit (Alpha Diagnostic International Inc.) according to manufacturer's protocol. The results were analyzed using SoftMaxpro 5.0, graph fitted to 4-pt fit curve and OD determined at 0.1.

Flow cytometry analysis of wild-type and glycoengineered PvDBPII

COS-7 cells were grown in 60 mm culture dishes containing one coverslip each. Cells were transfected using GeneJuice transfection reagent and 1 µg plasmid DNA of wild-type PvDBPII-GFP fusion (pEGFP-DBPII, kindly provided by Dr. John Adams, University of South Florida) or PvDBPII glycosylated variants (construct boundaries, DHKK…EVVT) (Figure S1) in pEGFP-N1 vector [23]. 48 h after transfection, the coverslips were transferred to a fresh culture dish for performing a COS-7-RBC binding assay described below. The remaining cells in the culture dish were lifted with 8% EDTA and washed with 1X PBS (0.1% BSA). The cells were then stained with rabbit anti-PvDBPII plasma for 1 h at 4^{o}C, followed by washing and labeling with goat anti-rabbit Texas Red (Southern Biotech) for 30 min. Antibody-labeled cells were washed and resuspended in PBS and analyzed by LSRII (Becton Dickinson). The data was analyzed by FlowJO 8.1 software (Tree Star Inc.)

COS-7 cells-RBC binding assay

To the coverslips generated above, 0.5% hematocrit blood was added and left for 1 h in a humidified chamber at 37^{o}C. The culture dish was then filled with complete DMEM, a glass stand was placed in the dish, and the submerged coverslip was flipped on the stand for 10 min to allow unbound RBC to settle by gravity. Coverslips were then washed with 1X PBS and fixed with 1% glutaraldehyde for 20 min. The coverslips were moved into 1X PBS and 10 fields were counted at 200X magnification using a fluroscence microscope. Transfected cells bearing 5 or more RBCs were counted as rosettes.

COS-7 cells-RBC antibody binding inhibition assay

COS-7 cells were cultured and transfected as described above. Coverslips were incubated with varying dilutions of final bleed mouse plasma or a 1:200 dilution of pre-bleed mouse plasma in incomplete DMEM for 1 h at 37^{o}C in a humidified chamber [23]. Coverslips were washed in 1X PBS and layered with 0.5% hematocrit blood for 1 h in a humidified chamber at 37^{o}C. The cells were then washed, fixed, and counted,

as described above. Percent inhibition was calculated at each plasma dilution as the percent decrease in RBC rosette count with plasma incubation relative to the rosette count without plasma using the formula [(percent rosette rate for no plasma - percent rosette rate for plasma dilution) / percent rosette rate for no plasma] x 100. For each immunization group, an inhibition curve (percent inhibition versus plasma dilution) was generated using GraphPad Prism version 5.02 (GraphPad Software) and the plasma dilution at which 50% inhibition was recorded (IC50) was determined by transforming the data to a log10 scale with fitted sigmoidal dose-response curves.

Yeast display antibody binding inhibition assay of PvDBPII and DARC-Fc

Yeast display of PvDBPII constructs was performed as per [39], using PvDBPII construct boundaries corresponding to the 2C6J PkDBPII crystallization construct [14]. In brief, a PvDBPII insert (amino acids, AIIN…PISQ) (Figure S1) was expressed as an Aga2p fusion protein at the yeast cell surface under the control of a galactose-inducible promoter in the episomally maintained yeast display plasmid pCTCON2. A linear PvDBPII insert with flanking regions homologous to pCTCON2 was designed using DNAWorks and synthesized by PCR assembly of oligonucleotides [69]. Nucleotides were codon-optimized for *S. cerevisiae* and all asparagines in potential N-linked glycosylation sites were mutated to glutamines. Viable plasmid was generated by homologous recombination following electroporation into EBY100 yeast as reported by Chao et al. [39]. Yeast clones were isolated by serial dilution on selective media and sequenced using the Zymo Research Yeast Plasmid Miniprep II kit. A yeast colony carrying the desired PvDBPII sequence was used for all further work. Induction of the Aga2p-PvDBPII fusion protein from the pCTCON2 vector and passaging of the transformed yeast were performed as described [39].

Binding and flow cytometry assays were modeled on the primary-secondary antibody protocol of Chao et al. [39]. In brief, 80,000 induced yeast per well were deposited in a 96-well plate, centrifuged for 5 minutes at 700 g, washed with PBSF (PBS + 0.1% bovine serum albumin) and incubated for 10 minutes with serial dilutions of immune plasma. The yeast cells were then washed with PBSF again and incubated with a 1:100 dilution of DARC-Fc (10 mg/L stock) in PBSF. After 10 mins incubation, the cells were washed with PBSF again and binding of the DARC-Fc protein was detected with a 1:250 dilution of phycoerythrin-conjugated goat anti-human polyclonal antibody preadsorbed against mouse and rabbit antibodies. DARC-Fc and DARC-Fc Y43F inactive mutant [17] were used as positive and negative

controls respectively in wells without immune plasma incubations. Antibody-labeled cells were washed and resuspended in PBS and analyzed by LSRII (Becton Dickinson). The data was analyzed by FlowJO 8.1 software (Tree Star Inc.). To calculate plasma inhibition, the population of PvDBPII expressing yeast was first determined by comparing uninduced to induced yeast after incubation with DARC-Fc and secondary antibody. The percent inhibition at each plasma dilution was then calculated as the percentage reduction in the shifted yeast population between the immune and pre-immune plasma. An inhibition curve and IC50 values were generated using GraphPad Prism as described above.

PCR-RFLP analysis to determine the DARC promoter type and FyA or FyB allele

The following protocol was adapted from [4]. Genomic DNA was extracted from 10 mls of whole blood using the QIAamp blood extraction kit following the manufacturer's directions (Qiagen). PCR was performed in 50 µl reactions containing 200 nM of the appropriate positive-strand and negative-strand primer (IDT DNA technologies); 200 µM each dNTP (New England Biolabs); 1.25 units of Taq DNA polymerase (25 and 50 µL reactions respectively; NEB); 1x polymerase buffer (NEB) and at least 200 ng of purified human genomic DNA. Two amplicons were used to characterize FY polymorphisms. The promoter fragment was characterized with (FYPup 5'-CTTCGGTAAAATCT CTACTTGCTGGAAAGC-3' and FYPdn 5'-CCATGGCACCGTTTGGTTCAGG-3') and the coding sequence with (FYup 5'-GACTCTTCCGGTGTAACTCTGATG-3' and FY851[-] 5'GGCCAAGACGGGCACCACAATG-3'). The thermocycling program used for both amplifications was 30 seconds at 95°C, 30 seconds at 58°C, and 30 seconds at 68°C for 30 to 45 cycles. The promoter-specific fragment was digested with StyI, and the coding-specific fragment digested with BanI for 2 h at 37°C. Restriction fragments were visualized after electrophoresis on 4% agarose (Lonza) gels stain. BanI digestion of the coding-specific PCR produces segments of 212 and 151 bp (FY*A allele) and/or 363 bp (Fy*B allele). StyI digestion of the Fy promoter-specific fragment produces segments of 144, 108, and 77 bp (erythrocyte active promoter) or 144, 108, 65 bp (erythrocyte silent promoter).

Entropy calculations

The amino acid variation at single multiple alignment position of a line-up of PvDBP sequences (Figure S2) was calculated using the formula:

Fractional Shannon entropy values $H(x) = -\sum_{i=1}^{20} \frac{P_i \log(P_i)}{0.05 \cdot \log(0.05)}$

where P_i represents the observed frequency of a residue type *i* in the aligned column [70]; the probability of a particular residue occurring at a given position is thus 0.05 (=1/20) for a uniform amino acid distribution, and the choice of logarithmic base is rendered irrelevant by the normalization process. The minimum positional entropy (0) occurs at perfectly conserved positions and the maximum positional entropy (4.32 = $\log_2$ 20) occurs at positions where all amino acids are observed with equal frequency.

ACKNOWLEDGMENTS

We thank Dr. Yves Durocher, Biotechnology Research Institute, National Research Council Canada for providing the pTT3 vector. We also thank Dr. Marion Avril and Jay Brazier for helpful discussion and assistance with flow cytometry analysis.

493 Reference List
494

495 1. Chitnis CE, Sharma A (2008) Targeting the Plasmodium vivax Duffy-binding protein. Trends
496 Parasitol 24: 29-34.

497 2. Miller LH, Mason SJ, Clyde DF, McGinniss MH (1976) The resistance factor to Plasmodium vivax
498 in blacks. The Duffy-blood-group genotype, FyFy. N Engl J Med 295: 302-304.

499 3. Tournamille C, Colin Y, Cartron JP, Le Van KC (1995) Disruption of a GATA motif in the Duffy
500 gene promoter abolishes erythroid gene expression in Duffy-negative individuals. Nat
501 Genet 10: 224-228.

502 4. Zimmerman PA, Woolley I, Masinde GL, Miller SM, McNamara DT et al. (1999) Emergence of
503 FY*A(null) in a Plasmodium vivax-endemic region of Papua New Guinea. Proc Natl Acad
504 Sci U S A 96: 13973-7.

505 5. Menard D, Barnadas C, Bouchier C, Henry-Halldin C, Gray LR et al. (2010) Plasmodium vivax
506 clinical malaria is commonly observed in Duffy-negative Malagasy people. Proc Natl
507 Acad Sci U S A 107: 5967-5971.

508 6. Wurtz N, Mint LK, Bogreau H, Pradines B, Rogier C et al. (2011) Vivax malaria in Mauritania
509 includes infection of a Duffy-negative individual. Malar J 10: 336.

510 7. Kasehagen LJ, Mueller I, Kiniboro B, Bockarie MJ, Reeder JC et al. (2007) Reduced Plasmodium
511 vivax erythrocyte infection in PNG Duffy-negative heterozygotes. PLoS ONE 2: e336.

512 8. King CL, Adams JH, Xianli J, Grimberg BT, McHenry AM et al. (2011) Fy(a)/Fy(b) antigen
513 polymorphism in human erythrocyte Duffy antigen affects susceptibility to Plasmodium
514 vivax malaria. Proc Natl Acad Sci U S A 108: 20113-20118.

515 9. Adams JH, Hudson DE, Torii M, Ward GE, Wellems TE et al. (1990) The Duffy receptor family of
516 Plasmodium knowlesi is located within the micronemes of invasive malaria merozoites.
517 Cell 63: 141-153.

518 10. Adams JH, Sim BK, Dolan SA, Fang X, Kaslow DC et al. (1992) A family of erythrocyte binding
519 proteins of malaria parasites. Proc Natl Acad Sci U S A 89: 7085-9.

520 11. Fang XD, Kaslow DC, Adams JH, Miller LH (1991) Cloning of the Plasmodium vivax Duffy
521 receptor. Mol Biochem Parasitol 44: 125-132.

522 12. Chitnis CE, Miller LH (1994) Identification of the erythrocyte binding domains of Plasmodium
523 vivax and Plasmodium knowlesi proteins involved in erythrocyte invasion. J Exp Med
524 180: 497-506.

525 13. Batchelor JD, Zahm JA, Tolia NH (2011) Dimerization of Plasmodium vivax DBP is induced upon
526 receptor binding and drives recognition of DARC. Nat Struct Mol Biol 18: 908-914.

527 14. Singh SK, Hora R, Belrhali H, Chitnis CE, Sharma A (2006) Structural basis for Duffy recognition
528 by the malaria parasite Duffy-binding-like domain. Nature 439: 741-744.

529 15. Choe H, Moore MJ, Owens CM, Wright PL, Vasilieva N et al. (2005) Sulphated tyrosines mediate
530 association of chemokines and Plasmodium vivax Duffy binding protein with the Duffy
531 antigen/receptor for chemokines (DARC). Mol Microbiol 55: 1413-1422.

16. Horuk R, Chitnis CE, Darbonne WC, Colby TJ, Rybicki A et al. (1993) A receptor for the malarial parasite Plasmodium vivax: the erythrocyte chemokine receptor. Science 261: 1182-4.

17. Hans D, Pattnaik P, Bhattacharyya A, Shakri AR, Yazdani SS et al. (2005) Mapping binding residues in the Plasmodium vivax domain that binds Duffy antigen during red cell invasion. Mol Microbiol 55: 1423-1434.

18. VanBuskirk KM, Sevova E, Adams JH (2004) Conserved residues in the Plasmodium vivax Duffy-binding protein ligand domain are critical for erythrocyte receptor recognition. Proc Natl Acad Sci U S A 101: 15754-15759.

19. Bolton MJ, Garry RF (2011) Sequence similarity between the erythrocyte binding domain 1 of the Plasmodium vivax Duffy binding protein and the V3 loop of HIV-1 strain MN reveals binding residues for the Duffy Antigen Receptor for Chemokines. Virol J 8: 45-54.

20. Chootong P, Ntumngia FB, VanBuskirk KM, Xainli J, Cole-Tobian JL et al. (2010) Mapping epitopes of the Plasmodium vivax Duffy binding protein with naturally acquired inhibitory antibodies. Infect Immun 78: 1089-1095.

21. Grimberg BT, Udomsangpetch R, Xainli J, McHenry A, Panichakul T et al. (2007) Plasmodium vivax invasion of human erythrocytes inhibited by antibodies directed against the Duffy binding protein. PLoS Med 4: e337.

22. King CL, Michon P, Shakri AR, Marcotty A, Stanisic D et al. (2008) Naturally acquired Duffy-binding protein-specific binding inhibitory antibodies confer protection from blood-stage Plasmodium vivax infection. Proc Natl Acad Sci U S A 105: 8363-8368.

23. Michon P, Fraser T, Adams JH (2000) Naturally acquired and vaccine-elicited antibodies block erythrocyte cytoadherence of the Plasmodium vivax Duffy binding protein. Infect Immun 68: 3164-3171.

24. Siddiqui AA, Xainli J, Schloegel J, Carias L, Ntumngia F et al. (2012) Fine Specificity of Plasmodium vivax Duffy Binding Protein Binding Engagement of the Duffy Antigen on Human Erythrocytes. Infect Immun 80: 2920-2928.

25. VanBuskirk KM, Cole-Tobian JL, Baisor M, Sevova ES, Bockarie M et al. (2004) Antigenic drift in the ligand domain of Plasmodium vivax duffy binding protein confers resistance to inhibitory antibodies. J Infect Dis 190: 1556-1562.

26. Dvorak JA, Miller LH, Whitehouse WC, Shiroishi T (1975) Invasion of erythrocytes by malaria merozoites. Science 187: 748-750.

27. Gilson PR, Crabb BS (2009) Morphology and kinetics of the three distinct phases of red blood cell invasion by Plasmodium falciparum merozoites. Int J Parasitol 39: 91-96.

28. Gosi P, Khusmith S, Khalambaheti T, Lanar DE, Schaecher KE et al. (2008) Polymorphism patterns in Duffy-binding protein among Thai Plasmodium vivax isolates. Malar J 7: 112.

29. Xainli J, Adams JH, King CL (2000) The erythrocyte binding motif of plasmodium vivax duffy binding protein is highly polymorphic and functionally conserved in isolates from Papua New Guinea. Mol Biochem Parasitol 111: 253-260.

30. Karlsson Hedestam GB, Fouchier RA, Phogat S, Burton DR, Sodroski J et al. (2008) The challenges of eliciting neutralizing antibodies to HIV-1 and to influenza virus. Nat Rev Microbiol 6: 143-155.

31. Garrity RR, Rimmelzwaan G, Minassian A, Tsai WP, Lin G et al. (1997) Refocusing neutralizing antibody response by targeted dampening of an immunodominant epitope. J Immunol 159: 279-289.

32. Pantophlet R, Wilson IA, Burton DR (2003) Hyperglycosylated mutants of human immunodeficiency virus (HIV) type 1 monomeric gp120 as novel antigens for HIV vaccine design. J Virol 77: 5889-5901.

33. Pantophlet R, Wilson IA, Burton DR (2004) Improved design of an antigen with enhanced specificity for the broadly HIV-neutralizing antibody b12. Protein Eng Des Sel 17: 749-758.

34. Selvarajah S, Puffer B, Pantophlet R, Law M, Doms RW et al. (2005) Comparing antigenicity and immunogenicity of engineered gp120. J Virol 79: 12148-12163.

35. Selvarajah S, Puffer BA, Lee FH, Zhu P, Li Y et al. (2008) Focused dampening of antibody response to the immunodominant variable loops by engineered soluble gp140. AIDS Res Hum Retroviruses 24: 301-314.

36. Samuelson J, Banerjee S, Magnelli P, Cui J, Kelleher DJ et al. (2005) The diversity of dolichol-linked precursors to Asn-linked glycans likely results from secondary loss of sets of glycosyltransferases. Proc Natl Acad Sci U S A 102: 1548-1553.

37. Petrescu AJ, Milac AL, Petrescu SM, Dwek RA, Wormald MR (2004) Statistical analysis of the protein environment of N-glycosylation sites: implications for occupancy, structure, and folding. Glycobiology 14: 103-114.

38. Mayer DC, Mu JB, Kaneko O, Duan J, Su XZ et al. (2004) Polymorphism in the Plasmodium falciparum erythrocyte-binding ligand JESEBL/EBA-181 alters its receptor specificity. Proc Natl Acad Sci U S A 101: 2518-2523.

39. Chao G, Lau WL, Hackel BJ, Sazinsky SL, Lippow SM et al. (2006) Isolating and engineering human antibodies using yeast surface display. Nat Protoc 1: 755-768.

40. Wang R, Smith JD, Kappe SH (2009) Advances and challenges in malaria vaccine development. Expert Rev Mol Med 11: e39.

41. Arevalo-Herrera M, Castellanos A, Yazdani SS, Shakri AR, Chitnis CE et al. (2005) Immunogenicity and protective efficacy of recombinant vaccine based on the receptor-binding domain of the Plasmodium vivax Duffy binding protein in Aotus monkeys. Am J Trop Med Hyg 73: 25-31.

42. Devi YS, Mukherjee P, Yazdani SS, Shakri AR, Mazumdar S et al. (2007) Immunogenicity of Plasmodium vivax combination subunit vaccine formulated with human compatible adjuvants in mice. Vaccine 25: 5166-5174.

43. Ntumngia FB, Adams JH (2012) Design and immunogenicity of a novel synthetic antigen based on the ligand domain of the Plasmodium vivax duffy binding protein. Clin Vaccine Immunol 19: 30-36.

44. Yazdani SS, Shakri AR, Mukherjee P, Baniwal SK, Chitnis CE (2004) Evaluation of immune responses elicited in mice against a recombinant malaria vaccine based on Plasmodium vivax Duffy binding protein. Vaccine 22: 3727-3737.

45. Wiley SR, Raman VS, Desbien A, Bailor HR, Bhardwaj R et al. (2011) Targeting TLRs expands the antibody repertoire in response to a malaria vaccine. Sci Transl Med 3: 93ra69.

614 46. Alexander S, Elder JH (1984) Carbohydrate dramatically influences immune reactivity of antisera
615 to viral glycoprotein antigens. Science 226: 1328-1330.

616 47. Wei X, Decker JM, Wang S, Hui H, Kappes JC et al. (2003) Antibody neutralization and escape
617 by HIV-1. Nature 422: 307-312.

618 48. Wyatt R, Kwong PD, Desjardins E, Sweet RW, Robinson J et al. (1998) The antigenic structure of
619 the HIV gp120 envelope glycoprotein. Nature 393: 705-711.

620 49. Back NK, Smit L, de Jong JJ, Keulen W, Schutten M et al. (1994) An N-glycan within the human
621 immunodeficiency virus type 1 gp120 V3 loop affects virus neutralization. Virology 199:
622 431-438.

623 50. Binley JM, Ban YE, Crooks ET, Eggink D, Osawa K et al. (2010) Role of complex carbohydrates
624 in human immunodeficiency virus type 1 infection and resistance to antibody
625 neutralization. J Virol 84: 5637-5655.

626 51. Bolmstedt A, Sjolander S, Hansen JE, Akerblom L, Hemming A et al. (1996) Influence of N-linked
627 glycans in V4-V5 region of human immunodeficiency virus type 1 glycoprotein gp160 on
628 induction of a virus-neutralizing humoral response. J Acquir Immune Defic Syndr Hum
629 Retrovirol 12: 213-220.

630 52. McCaffrey RA, Saunders C, Hensel M, Stamatatos L (2004) N-linked glycosylation of the V3 loop
631 and the immunologically silent face of gp120 protects human immunodeficiency virus
632 type 1 SF162 from neutralization by anti-gp120 and anti-gp41 antibodies. J Virol 78:
633 3279-3295.

634 53. Calarese DA, Scanlan CN, Zwick MB, Deechongkit S, Mimura Y et al. (2003) Antibody domain
635 exchange is an immunological solution to carbohydrate cluster recognition. Science 300:
636 2065-2071.

637 54. McLellan JS, Pancera M, Carrico C, Gorman J, Julien JP et al. (2011) Structure of HIV-1 gp120
638 V1/V2 domain with broadly neutralizing antibody PG9. Nature 480: 336-343.

639 55. Pejchal R, Doores KJ, Walker LM, Khayat R, Huang PS et al. (2011) A potent and broad
640 neutralizing antibody recognizes and penetrates the HIV glycan shield. Science 334:
641 1097-1103.

642 56. Walker LM, Simek MD, Priddy F, Gach JS, Wagner D et al. (2010) A limited number of antibody
643 specificities mediate broad and potent serum neutralization in selected HIV-1 infected
644 individuals. PLoS Pathog 6: e1001028.

645 57. Pancera M, McLellan JS, Wu X, Zhu J, Changela A et al. (2010) Crystal structure of PG16 and
646 chimeric dissection with somatically related PG9: structure-function analysis of two
647 quaternary-specific antibodies that effectively neutralize HIV-1. J Virol 84: 8098-8110.

648 58. Pejchal R, Walker LM, Stanfield RL, Phogat SK, Koff WC et al. (2010) Structure and function of
649 broadly reactive antibody PG16 reveal an H3 subdomain that mediates potent
650 neutralization of HIV-1. Proc Natl Acad Sci U S A 107: 11483-11488.

651 59. Schief WR, Ban YE, Stamatatos L (2009) Challenges for structure-based HIV vaccine design.
652 Curr Opin HIV AIDS 4: 431-440.

60. Drummer HE, Jackson DC, Brown LE (1993) Modulation of CD4+ T-cell recognition of influenza hemagglutinin by carbohydrate side chains located outside a T-cell determinant. Virology 192: 282-289.

61. Jackson DC, Drummer HE, Urge L, Otvos L, Jr., Brown LE (1994) Glycosylation of a synthetic peptide representing a T-cell determinant of influenza virus hemagglutinin results in loss of recognition by CD4+ T-cell clones. Virology 199: 422-430.

62. Srivastava A, Gangnard S, Round A, Dechavanne S, Juillerat A et al. (2010) Full-length extracellular region of the var2CSA variant of PfEMP1 is required for specific, high-affinity binding to CSA. Proc Natl Acad Sci U S A 107: 4884-4889.

63. Rohl CA, Strauss CE, Misura KM, Baker D (2004) Protein structure prediction using Rosetta. Methods Enzymol 383: 66-93.

64. Guex N, Peitsch MC (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis 18: 2714-2723.

65. Shakin-Eshleman SH, Spitalnik SL, Kasturi L (1996) The amino acid at the X position of an Asn-X-Ser sequon is an important determinant of N-linked core-glycosylation efficiency. J Biol Chem 271: 6363-6366.

66. Go EP, Irungu J, Zhang Y, Dalpathado DS, Liao HX et al. (2008) Glycosylation site-specific analysis of HIV envelope proteins (JR-FL and CON-S) reveals major differences in glycosylation site occupancy, glycoform profiles, and antigenic epitopes' accessibility. J Proteome Res 7: 1660-1674.

67. Sellhorn G, Caldwell Z, Mineart C, Stamatatos L (2009) Improving the expression of recombinant soluble HIV Envelope glycoproteins using pseudo-stable transient transfection. Vaccine 28: 430-436.

68. Davis HL, Weeratna R, Waldschmidt TJ, Tygrett L, Schorr J et al. (1998) CpG DNA is a potent enhancer of specific immunity in mice immunized with recombinant hepatitis B surface antigen. J Immunol 160: 870-876.

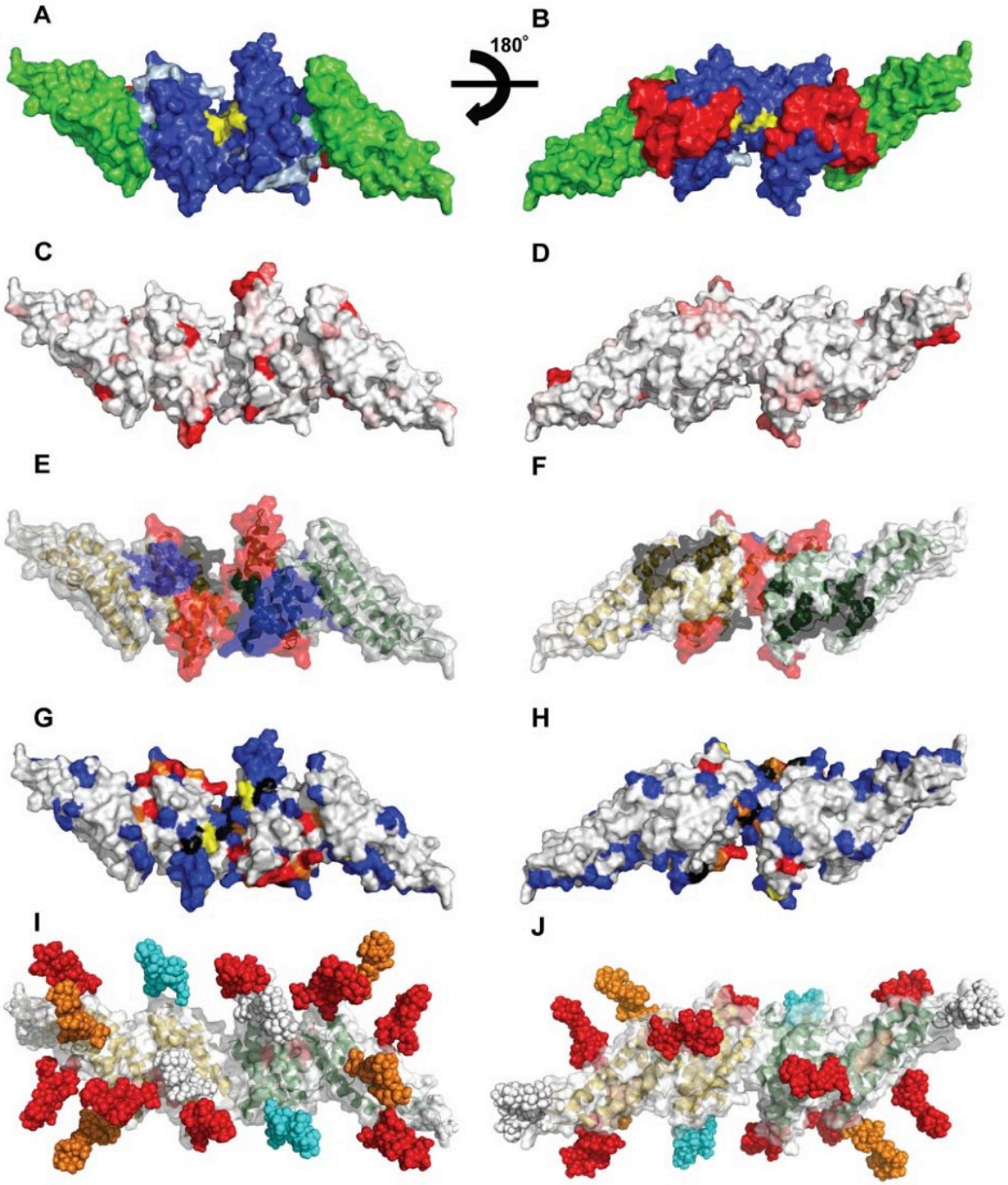
69. Azoitei ML, Correia BE, Ban YE, Carrico C, Kalyuzhniy O et al. (2011) Computation-guided backbone grafting of a discontinuous motif onto a protein scaffold. Science 334: 373-376.

70. Shannon CE (1948) A Mathematical Theory of Communication. Bell Sys Tech J 3: 379-423.

71. Hodder AN, Czabotar PE, Uboldi AD, Clarke OB, Lin CS et al. (2012) Insights into duffy binding-like domains through the crystal structure and function of the merozoite surface protein MSPDBL2 from P. falciparum. J Biol Chem 287: 32922-32939.

Figure 1. Design of surface re-engineered PvDBPII recombinant proteins containing additional N-linked glycan residues. (A-B) Subdomains and DARC binding models. Subdomain 1(red), subdomain 2 (blue), and subdomain 3 (green). Critical binding residues for model 1 are colored light blue and for model 2 are colored yellow. (C-D) Fractional Shannon entropy values [70] from 0.000 (white) to 0.243 (red) for sequence polymorphism over the PvDBPII surface as compared with maximally entropic distribution over

691 all amino acids. (E-F) Epitopes recognized by blocking antibodies [20]; black (low inhibitory), blue
692 (medium inhibitory), red (high inhibitory). (G-H) Meta-analysis of mutations that reduce or do not affect the
693 PvDBP-DARC interaction: blue residues (no effect); yellow residues (minor); orange residues (moderate);
694 red residues (major); black residues, differences between studies (see Table S1). (I-J) Location of
695 engineered N-glycosylation sites modeled as high mannose forms; white (wild type), cyan (STBP glycan),
696 Orange (P1 and Max), red (Max). All images modeled in PyMol on the PvDBPII dimer structure [13],
697 viewing opposite ends of the dimeric two-fold axis for subpanels I and II; missing density in the crystal
698 structure has been left missing, though it contains polymorphic, inhibitory, antibody recognition and
699 glycan-bearing sites. PvDBPII monomers are colored yellow and green in panels E, F, I, and J.



700

701 **Figure 2. Summary of PvDBP polymorphism, inhibitory epitopes, and residues impacting DARC**
702 **binding.** The sequence of the solved Sal1strain PvDBP variant crystal structure [13] is shown.
703 Polymorphic amino acids from 129 PvDBP sequences [28,29] are listed below the Sal1 sequence. Alpha
704 helices in the PvDBP crystal structure [13] are indicated by "h" and labeled helix 1a to helix 9 according to

convention [71]. Circles above the line-up indicate important residues. N-glycosylation sites are numbered according to Table 1 and colored green (wild type), blue (STBP glycan), orange (P1 and Max), and red (Max). Dimer interface [13] – black circles; polymorphism – light grey (rare, <10% of sequences), dark grey (> 10% of sequences); mutations that effect DARC binding from this study and others [13,17–19], are colored blue (no effect), yellow with black shadowing (minor), orange (moderate), red (major), and black, differences between studies (see Table S1); linear epitopes targeted by inhibitory antibodies [20] – black or grey shading (low inhibitory), blue shading (medium inhibitory), red shading (high inhibitory).

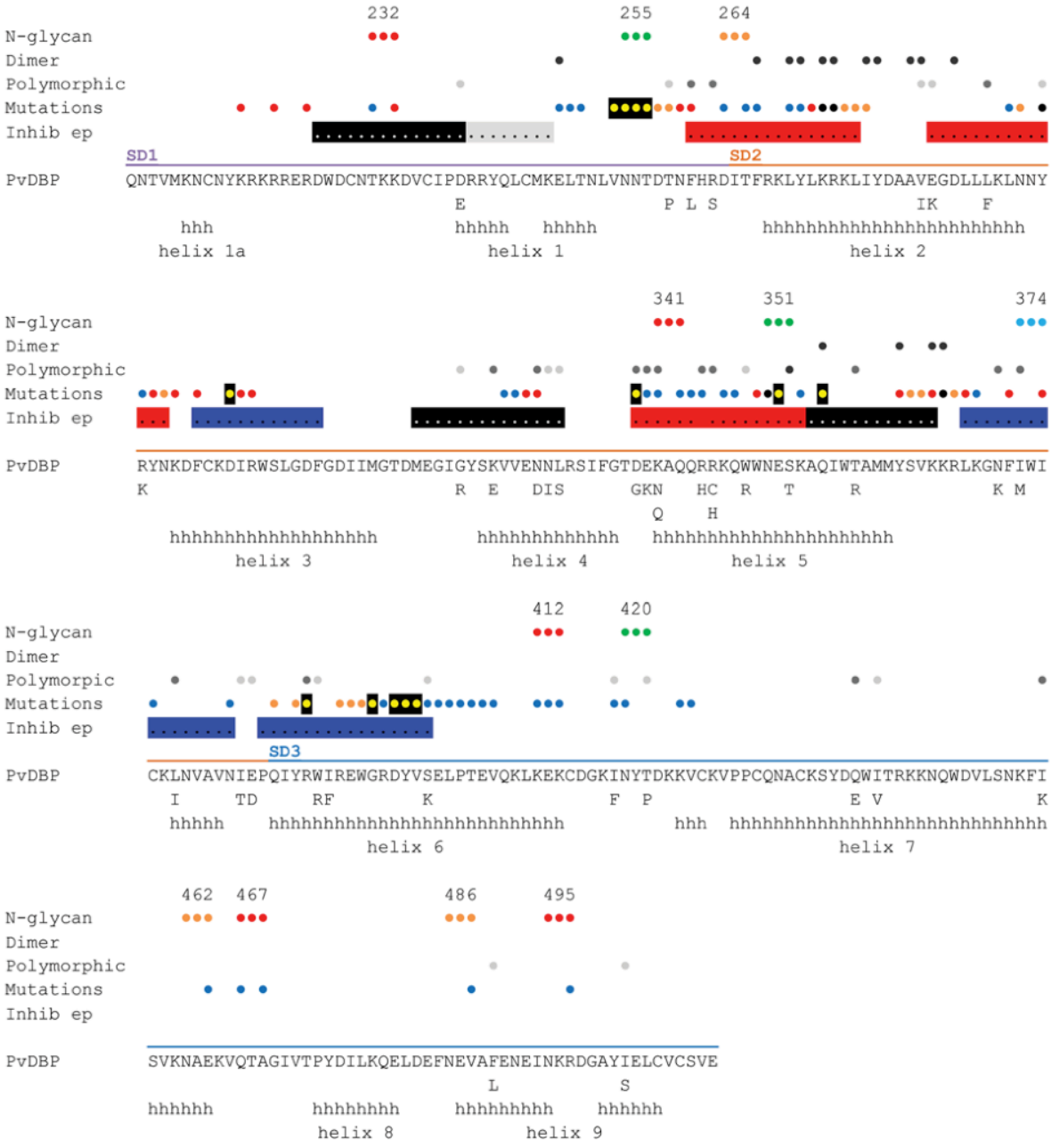


Figure 3. Expression and purification of surface re-engineered PvDBPII recombinant proteins containing additional N-linked glycan residues. (A) 2 µg of PvDBPII wild type and DBPII glycosylation variants were run on SDS-PAGE gel and stained with GelCode Blue reagent. A ladder effect is seen consistent with increasing number of N-glycosylation sites present in the STBP glycan, P1 and Max hyperglycosylated variants compared to PvDBPII wild-type. (B) Western blot of 1 µg of PvDBPII and DBPII glycosylation variants probed with anti-His antibody. (C) PvDBPII wild type and DBPII glycosylation

variants were either untreated (-) or digested (+) with N-glycosidase PNGaseF, run on SDS-PAGE gel and probed with anti-His antibody, except for Max variant which was probed with anti-PvDBPII serum. The P1 lanes were run separate from the other samples. Molecular mass is shown on the left.

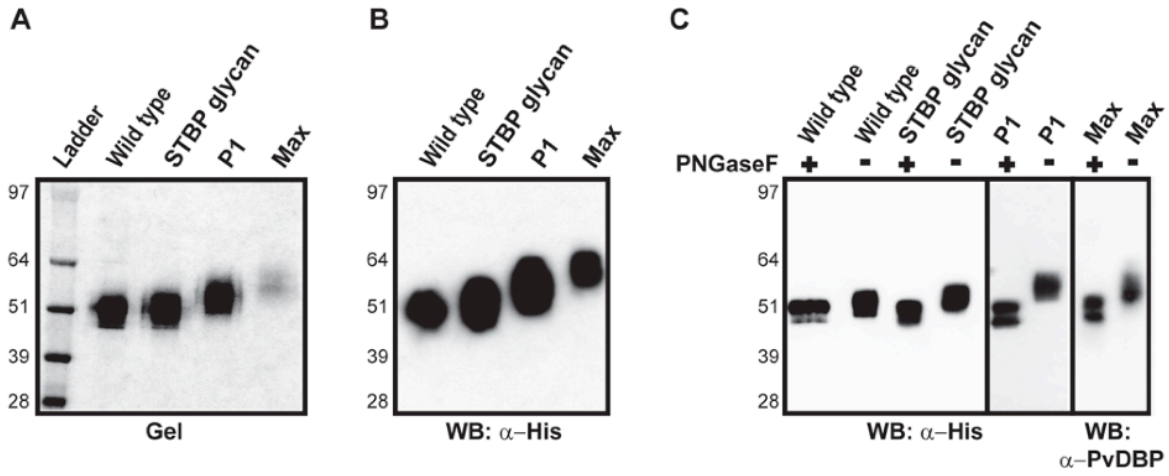
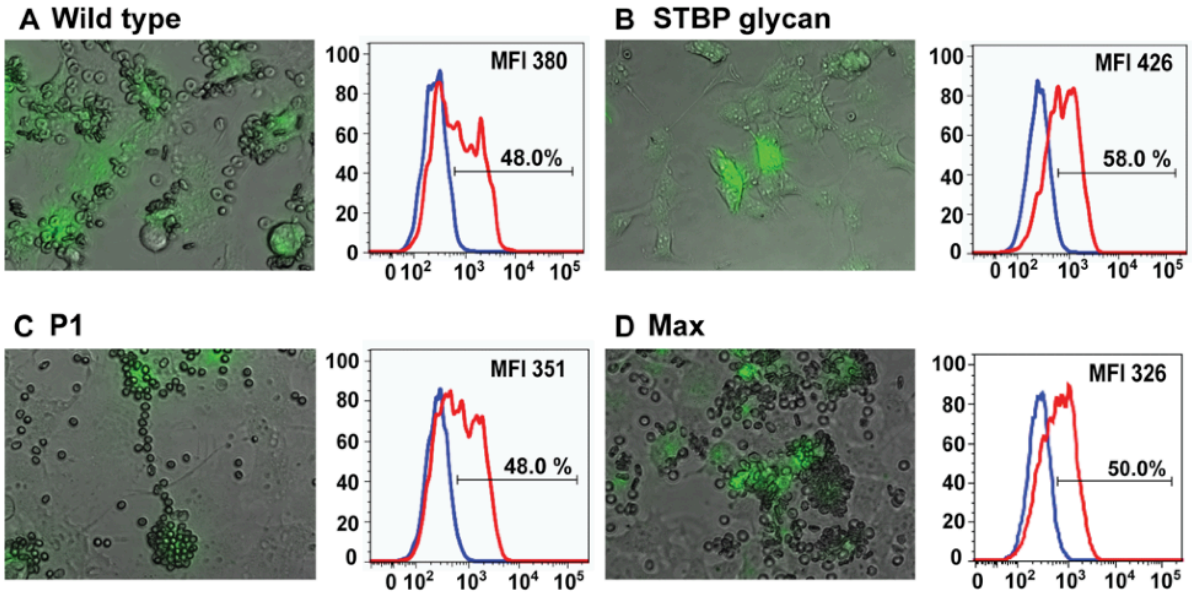
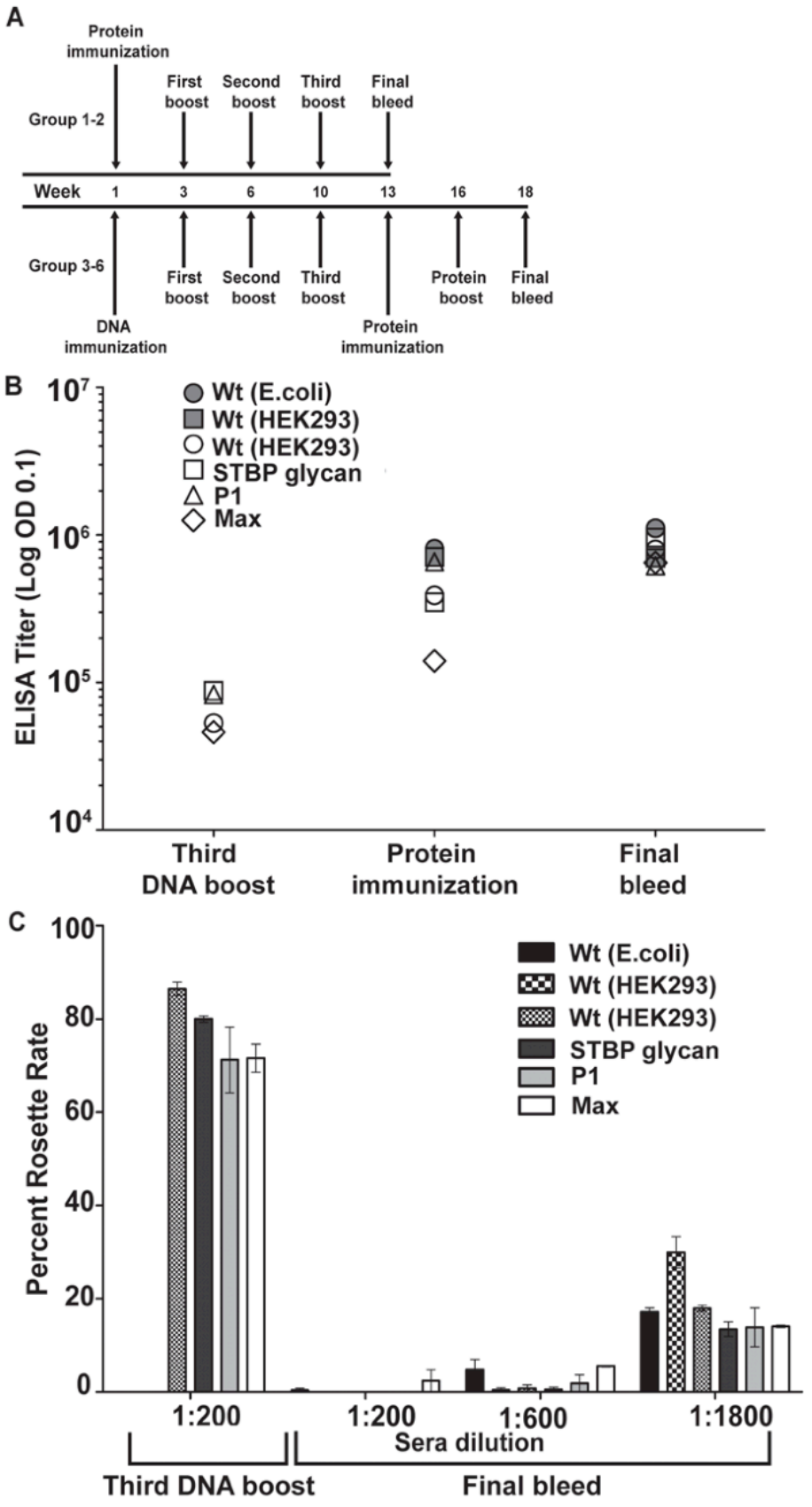


Figure 4. Functional characterization of hyperglycosylated PvDBPII variants by COS-7 cell-RBC binding assay. PvDBPII wild type (A), STBP glycan mutant (B), P1 (C) and Max (D) DBPII glycosylation variants were expressed as a GFP fusion protein in COS-7 cells and incubated with 0.5% hematocrit erythrocytes. GFP-positive transfected cells with five or more RBC (rosettes) were considered positive for binding. The histogram shows the percentage surface labeling when GFP-positive transfected cells were probed with pre-immune (blue) or anti-PvDBPII plasma (red). The gating strategy for GFP-positive and anti-PvDBPII positive is shown in Figure S3. The percentage signifies GFP-positive cells that are labeled with immune plasma, indicating surface expression of PvDBPII and DBPII glycosylation variants. The shift in mean fluorescence intensities between pre-immune and immune plasma is indicated as MFI.



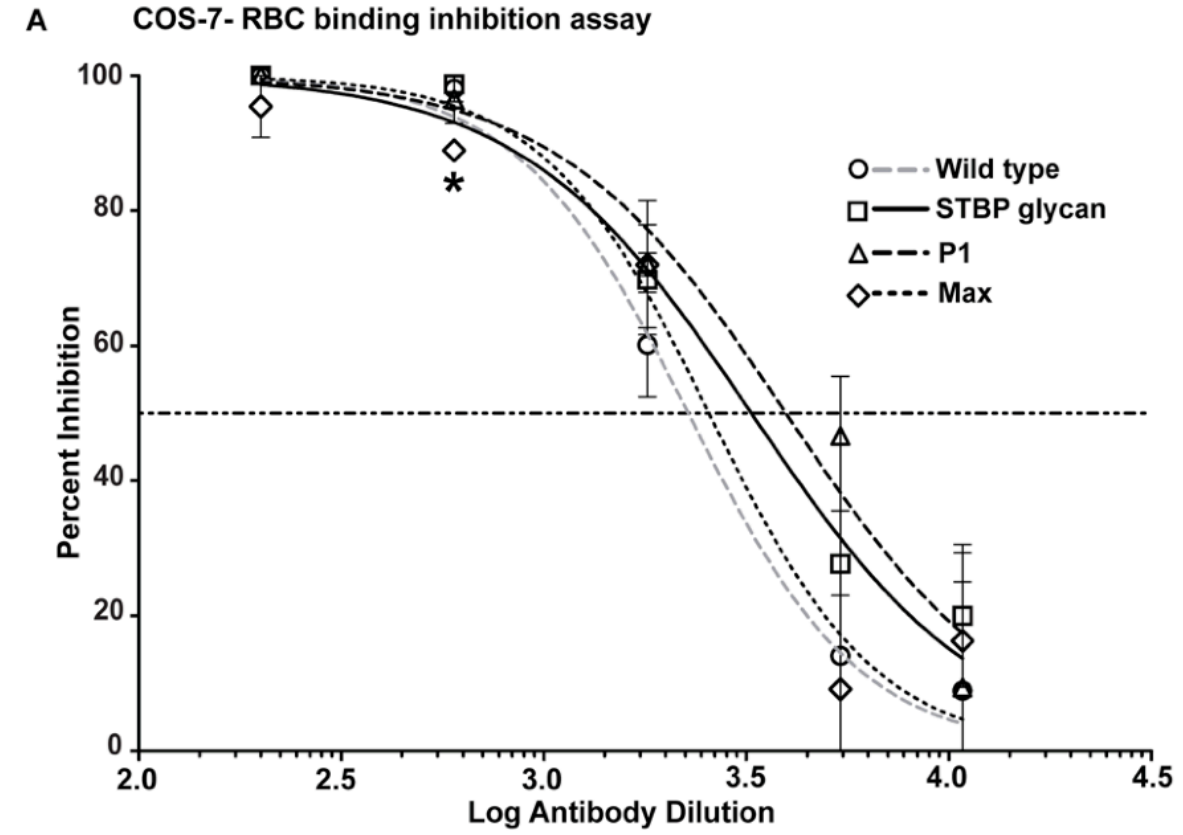
732

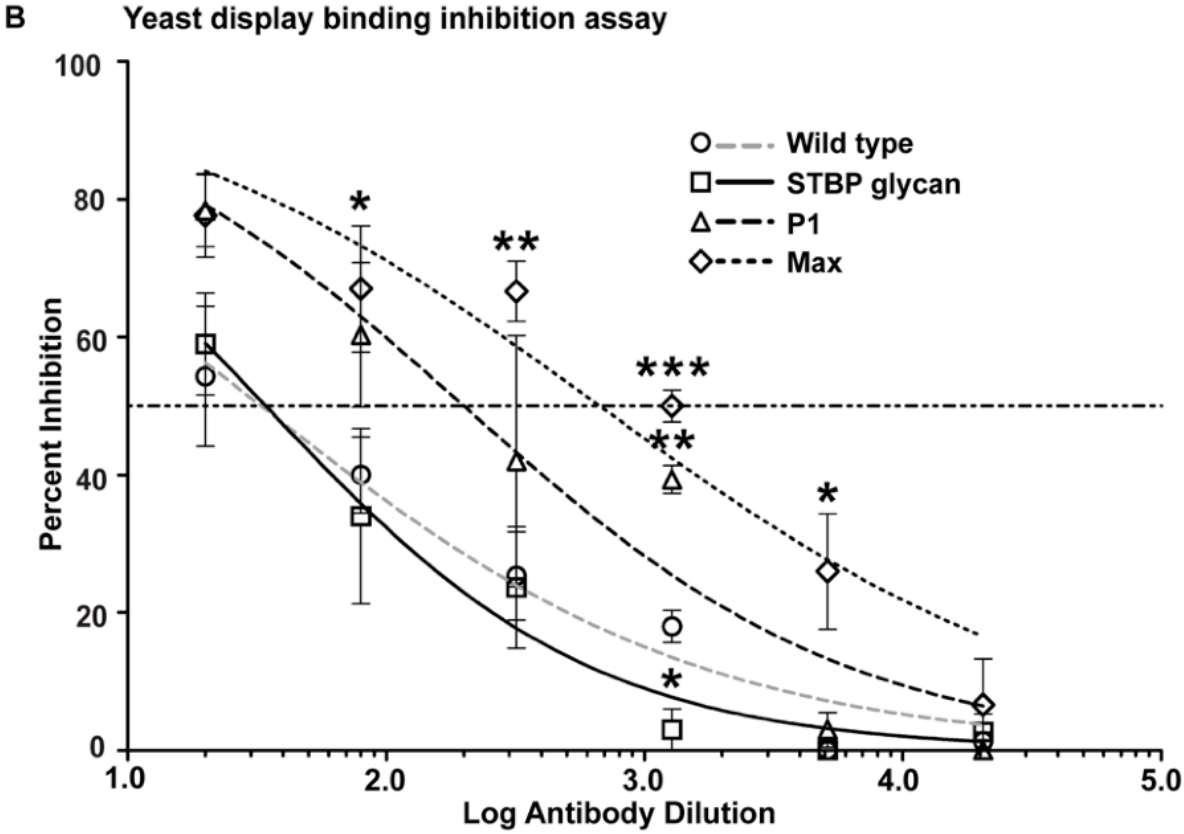
733 **Figure 5. Immunogenic characterization of PvDBPII and its hyperglycosylated variants**. (A)
734 Schematic representation of immunization scheme for PvDBPII wild type and DBPII glycosylation variants.
735 Group 1 mice received PvDBPII wild type generated in *E. coli*, Group 2 mice received PvDBPII wild type
736 generated in HEK-293F cells. Mice in groups 3- 6 received a combination of DNA and homologous
737 protein; group 3, PvDBPII wildtype; group 4, STBP glycan; group 5, P1; group 6, Max hyperglycosylated
738 variant. (B) The ELISA end-point antibody titers are shown for each of the immunization groups after the
739 third DNA boost, the first protein immunization, and the second protein boost (final bleed). (C) The
740 inhibitory activity of vaccine plasma in a COS-7-RBC binding assay. COS-7 cells expressing PvDBPII wild
741 type protein were pre-incubated with varying dilutions of PvDBPII wild type and DBPII glycosylation
742 variant immune plasma obtained after the third DNA boost or the second protein immunization (final
743 bleed). The values are an average of duplicate experiments.


A
Protein immunization
First boost
Second boost
Third boost
Final bleed
Group 1-2
Week
1
3
6
10
13
16
18
Group 3-6
First boost
Second boost
Third boost
Protein boost
Final bleed
DNA immunization
Protein immunization
B
Wt (E.coli)
Wt (HEK293)
Wt (HEK293)
STBP glycan
P1
Max
ELISA Titer (Log OD 0.1)
Third DNA boost
Protein immunization
Final bleed
C
Wt (E.coli)
Wt (HEK293)
Wt (HEK293)
STBP glycan
P1
Max
Percent Rosette Rate
1:200
1:200
1:600
1:1800
Sera dilution
Third DNA boost
Final bleed

744

745 **Figure 6. Inhibition of PvDBPII binding to DARC in different assay platforms.** (A) Inhibition of
746 PvDBPII – DARC interaction by immune plasma in the COS-7 cell**–RBC** binding inhibition assay. COS-7
747 cells expressing wild-type PvDBPII were incubated with varying dilutions of immune plasma obtained
748 after the final protein boost. The values are an average of two experiments. (B) Inhibition of PvDBPII –
749 DARC interaction by immune plasma in the yeast display binding inhibition assay. Yeast cells expressing
750 wild-type PvDBPII on the surface were incubated with varying dilutions of immune plasma obtained after
751 final protein boost and then probed with recombinant DARC-Fc. Values are an average of three
752 experiments, standard deviation is shown as error bars. Student's t test, p value < 0.001 is indicated by
753 ***, p value from 0.001 to 0.01 is shown by ** and 0.01 to 0.05 by *.


A
COS-7- RBC binding inhibition assay
Percent Inhibition
Log Antibody Dilution
Wild type
STBP glycan
P1
Max
100
80
60
40
20
0
2.0
2.5
3.0
3.5
4.0
4.5
*


B
Yeast display binding inhibition assay
Percent Inhibition
Log Antibody Dilution
Wild type
STBP glycan
P1
Max
100
80
60
40
20
0
1.0
2.0
3.0
4.0
5.0
*
**
***
**
*
*

754

Figure 7. PvDBP – DARC interaction. DARC, a seven transmembrane chemokine receptor on erythrocytes, has been shown to bind to PvDBP [16]. Binding has been mapped to the N-terminal 65 amino acids (illustrated as a red tail). The COS-7-RBC cytoadherance assay is based upon a multivalent interaction between PvDBPII present on surface of COS-7 cells and DARC expressed by RBC. In the yeast-PvDBPII display assay, PvDBPII present on the yeast surface interacts with dimeric recombinant DARC-Fc recombinant protein (N-terminal 65 mer region). The two assays offer different platforms to reveal inhibitory effects of antibodies using a potentially higher affinity, multimer-multimer interaction (COS-7 format) or a lower affinity multimer-dimer interaction (yeast display). Bottom of figure, polymorphisms in DARC that generate FyA/FyA^{null} genotype are associated with half of the number of surface DARC protein and lower *P. vivax* infection [4].

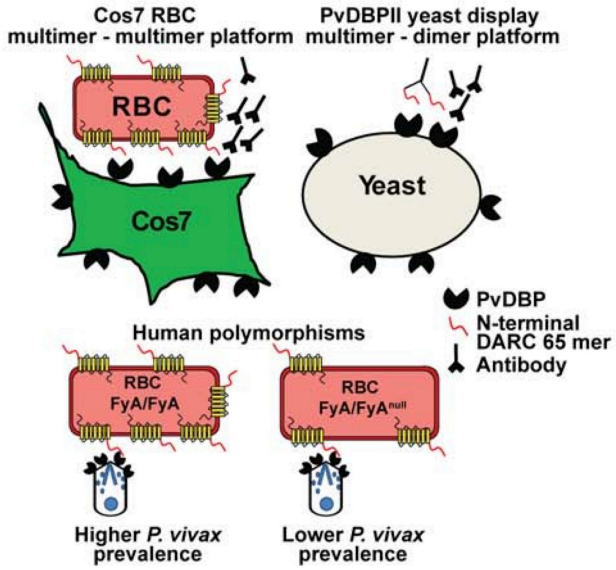
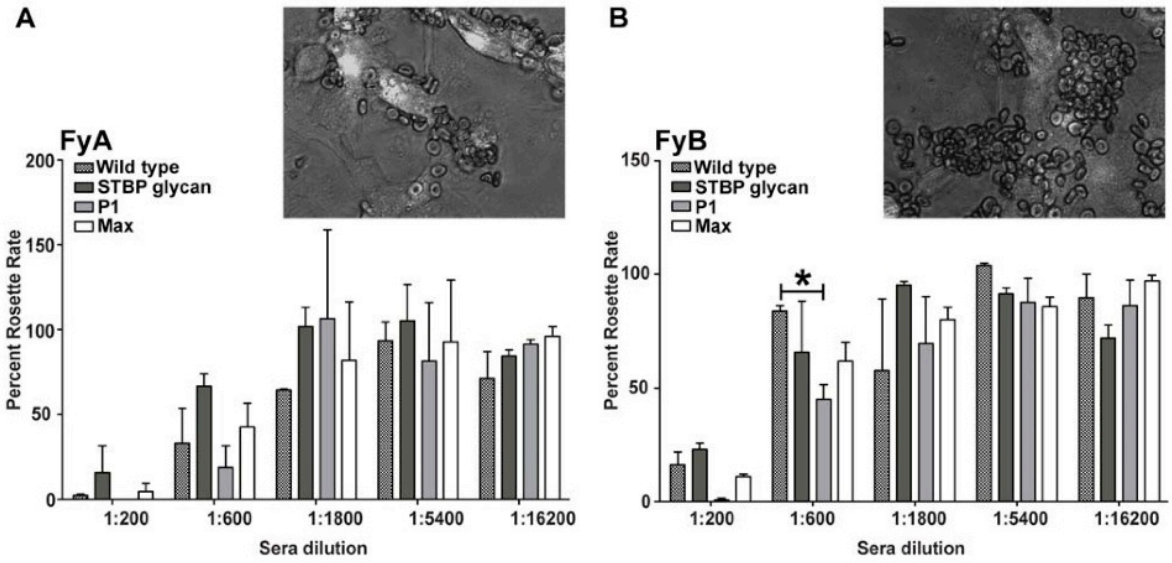


Figure 8. Effect of DARC phenotype on antibody binding inhibitory activity. COS-7 cells expressing PvDBPII as a GFP fusion protein were incubated with immune plasma and then RBC expressing the FyA (A) or FyB (B) Duffy blood group antigen were added. The inset shows that FyA RBCs (A) gave smaller rosettes than FyB RBCs (B) in the COS-7 cell**–RBC** binding assay. Statistical testing and p values as explained in Figure 6B.



771

772 **Figure S1. Sequence of PvDBPII wild type and glycoengineered variants.** N-glycan sites are
773 highlighted in yellow. The PvDBP sequence indicates the amino acid boundaries for the DNA and protein
774 immunization constructs. The PvDBP construct boundaries for COS-7 assay are underlined and those for
775 yeast display are shown in red.

Wild type

D H K K T I S S A I I N H A F L Q N T V M K N C N Y K R K R R E R D W D C N T
K K D V C I P D R R Y Q L C M K E L T N L V N N T D T N F H R D I T F R K L Y
L K R K L I Y D A A V E G D L L L K L N N Y R Y N K D F C K D I R W S L G D F
G D I I M G T D M E G I G Y S K V V E N N L R S I F G T D E K A Q Q R R K Q W
W N E S K A Q I W T A M M Y S V K K R L K G N F I W I C K L N V A V N I E P Q
I Y R W I R E W G R D Y V S E L P T E V Q K L K E K C D G K I N Y T D K K V C
K V P P C Q N A C K S Y D Q W I T R K K N Q W D V L S N K F I S V K N A E K V
Q T A G I V T P Y D I L K Q E L D E F N E V A F E N E I N K R D G A Y I E L C
V C S V E E A K K N T Q E V V T N V D N A A K S Q A T N S N P I S Q P V D S S
K A E K V P

STBP glycan

D H K K T I S S A I I N H A F L Q N T V M K N C N Y K R K R R E R D W D C N T
K K D V C I P D R R Y Q L C M K E L T N L V N N T D T N F H R D I T F R K L Y
L K R K L I Y D A A V E G D L L L K L N N Y R Y N K D F C K D I R W S L G D F
G D I I M G T D M E G I G Y S K V V E N N L R S I F G T D E K A Q Q R R K Q W
W N E S K A Q I W T A M M Y S V K K R L K G N F N W T C K L N V A V N I E P Q
I Y R W I R E W G R D Y V S E L P T E V Q K L K E K C D G K I N Y T D K K V C
K V P P C Q N A C K S Y D Q W I T R K K N Q W D V L S N K F I S V K N A E K V
Q T A G I V T P Y D I L K Q E L D E F N E V A F E N E I N K R D G A Y I E L C
V C S V E E A K K N T Q E V V T N V D N A A K S Q A T N S N P I S Q P V D S S
K A E K V P

P1

D H K K T I S S A I I N H A F L Q N T V M K N C N Y K R K R R E R D W D C N T
K K D V C I P D R R Y Q L C M K E L T N L V N N T D T N F H R N I T F R K L Y
L K R K L I Y D A A V E G D L L L K L N N Y R Y N K D F C K D I R W S L G D F
G D I I M G T D M E G I G Y S K V V E N N L R S I F G T D E K A Q Q R R K Q W
W N E S K A Q I W T A M M Y S V K K R L K G N F I W I C K L N V A V N I E P Q
I Y R W I R E W G R D Y V S E L P T E V Q K L K E K C D G K I N Y T D K K V C
K V P P C Q N A C K S Y D Q W I T R K K N Q W D V L S N K F I S V K N A S K V
Q T A G I V T P Y D I L K Q E L D E F N E T A F E N E I N K R D G A Y I E L C
V C S V E E A K K N T Q E V V T N V D N A A K S Q A T N S N P I S Q P V D S S
K A E K V P

Max

D H K K T I S S A I I N H A F L Q N T V M K N C N Y K R K R R E R D W D C N N
K S D V C I P D R R Y Q L C M K E L T N L V N N T D T N F H R N I T F R K L Y
L K R K L I Y D A A V E G D L L L K L N N Y R Y N K D F C K D I R W S L G D F
G D I I M G T D M E G I G Y S K V V E N N L R S I F G T D E N A S Q R R K Q W
W N E S K A Q I W T A M M Y S V K K R L K G N F I W I C K L N V A V N I E P Q
I Y R W I R E W G R D Y V S E L P T E V Q K L N E S C D G K I N Y T D K K V C
K V P P C Q N A C K S Y D Q W I T R K K N Q W D V L S N K F I S V K N A S K V
N T T G I V T P Y D I L K Q E L D E F N E T A F E N E I N K S D G A Y I E L C
V C S V E E A K K N T Q E V V T N V D N A A K S Q A T N S N P I S Q P V D S S
K A E K V P

776

777 **Figure S2. Alignment of PvDBP sequences.** A multiple alignment was generated by Clustal 2.1 from
778 129 PvDBP sequences [28,29] and used to generate the entropy scores at each amino acid position in
779 the line-up. *, fully conserved residue; : strong conservation; . weak conservation.

```
gi|11120233|gb|AAG30847.1|          DHKKTISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|12407747|gb|AAG53634.1|          DHKKTISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|18766659|gb|AAL79050.1|AF46      ---------------------------------------------DRRY
gi|18766819|gb|AAL79130.1|AF46      ---------------------------------------------DRRY
gi|63146462|gb|AAY34059.1|          ---------------------------------------------DRRY
gi|63146616|gb|AAY34136.1|          ---------------------------------------------DRRY
gi|11138983|gb|AAG31571.1|          DHKKTISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|73697858|gb|AAZ81532.1|          DHKKTISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|73697846|gb|AAZ81526.1|          DHKKTISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|12004297|gb|AAG43989.1|          -------------------------------ERDWDCNTKKDVCIPDRRY
gi|12004299|gb|AAG43990.1|          -------------------------------ERDWDCNTKKDVCIPDRRY
gi|73697866|gb|AAZ81536.1|          DHKKTISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|73697848|gb|AAZ81527.1|          DHKKTISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|73697856|gb|AAZ81531.1|          DHKKTISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|73697862|gb|AAZ81534.1|          DHKKTISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|160276|gb|AAA63423.1|            DHKKTISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|73697844|gb|AAZ81525.1|          DHKKTISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|339961368|pdb|3RRC|A             ---------------MGNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|339961369|pdb|3RRC|B             ---------------MGNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|255914751|dbj|BAH96577.1|        -------------------------------ERDWDCNTKKDVCIPDRRY
gi|224926624|gb|ACN69874.1|         ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926626|gb|ACN69875.1|         ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926628|gb|ACN69876.1|         ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926630|gb|ACN69877.1|         ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926632|gb|ACN69878.1|         ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926634|gb|ACN69879.1|         ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926636|gb|ACN69880.1|         ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926638|gb|ACN69881.1|         ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926640|gb|ACN69882.1|         ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926642|gb|ACN69883.1|         ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDGCIPERRY
gi|224926644|gb|ACN69884.1|         ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926646|gb|ACN69885.1|         ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926648|gb|ACN69886.1|         ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926650|gb|ACN69887.1|         ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926652|gb|ACN69888.1|         ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926654|gb|ACN69889.1|         ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926656|gb|ACN69890.1|         ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926658|gb|ACN69891.1|         ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926660|gb|ACN69892.1|         ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926662|gb|ACN69893.1|         ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926664|gb|ACN69894.1|         ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926666|gb|ACN69895.1|         ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926668|gb|ACN69896.1|         ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926670|gb|ACN69897.1|         ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926672|gb|ACN69898.1|         ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926674|gb|ACN69899.1|         ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926676|gb|ACN69900.1|         ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926678|gb|ACN69901.1|         ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926680|gb|ACN69902.1|         ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926682|gb|ACN69903.1|         ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926684|gb|ACN69904.1|         ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926686|gb|ACN69905.1|         ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926688|gb|ACN69906.1|         ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926690|gb|ACN69907.1|         ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926692|gb|ACN69908.1|         ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926694|gb|ACN69909.1|         ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926696|gb|ACN69910.1|         ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926698|gb|ACN69911.1|         ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926700|gb|ACN69912.1|         ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
```

```
gi|224926702|gb|ACN69913.1|        ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926704|gb|ACN69914.1|        ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926706|gb|ACN69915.1|        ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926708|gb|ACN69916.1|        ----TISSAIINHALLQNTVMKNCNYKRKRRERDWGCNTKKDVCIPDRRY
gi|224926710|gb|ACN69917.1|        ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926714|gb|ACN69918.1|        ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926716|gb|ACN69919.1|        ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926718|gb|ACN69920.1|        ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926720|gb|ACN69921.1|        ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926722|gb|ACN69922.1|        ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926724|gb|ACN69923.1|        ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926726|gb|ACN69924.1|        ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926728|gb|ACN69925.1|        ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926730|gb|ACN69926.1|        ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926732|gb|ACN69927.1|        ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926734|gb|ACN69928.1|        ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926736|gb|ACN69929.1|        ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926738|gb|ACN69930.1|        ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926740|gb|ACN69931.1|        ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926742|gb|ACN69932.1|        ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926744|gb|ACN69933.1|        ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926746|gb|ACN69934.1|        ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926748|gb|ACN69935.1|        ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926750|gb|ACN69936.1|        ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926752|gb|ACN69937.1|        ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926754|gb|ACN69938.1|        ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926756|gb|ACN69939.1|        ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926758|gb|ACN69940.1|        ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926760|gb|ACN69941.1|        ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926762|gb|ACN69942.1|        ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926764|gb|ACN69943.1|        ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926766|gb|ACN69944.1|        ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926768|gb|ACN69945.1|        ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926770|gb|ACN69946.1|        ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926772|gb|ACN69947.1|        ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926774|gb|ACN69948.1|        ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926776|gb|ACN69949.1|        ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926778|gb|ACN69950.1|        ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926780|gb|ACN69951.1|        ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926782|gb|ACN69952.1|        ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926784|gb|ACN69953.1|        ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926786|gb|ACN69954.1|        ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926788|gb|ACN69955.1|        ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926790|gb|ACN69956.1|        ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926792|gb|ACN69957.1|        ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926794|gb|ACN69958.1|        ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926796|gb|ACN69959.1|        ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926798|gb|ACN69960.1|        ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926800|gb|ACN69961.1|        ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926802|gb|ACN69962.1|        ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926804|gb|ACN69963.1|        ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926806|gb|ACN69964.1|        ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926808|gb|ACN69965.1|        ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926810|gb|ACN69966.1|        ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926812|gb|ACN69967.1|        ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926814|gb|ACN69968.1|        ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926816|gb|ACN69969.1|        ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926818|gb|ACN69970.1|        ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|224926820|gb|ACN69971.1|        ----TISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
```

```
gi|213868281|gb|ACJ54187.1|         DHKKTISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|213868283|gb|ACJ54188.1|         DHKKTISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|213868285|gb|ACJ54189.1|         DHKKTISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|213868287|gb|ACJ54190.1|         DHKKTISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|213868289|gb|ACJ54191.1|         DHKKTISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|213868291|gb|ACJ54192.1|         DHKKTISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|213868293|gb|ACJ54193.1|         DHKKTISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|213868295|gb|ACJ54194.1|         DHKKTISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|213868297|gb|ACJ54195.1|         DHKKTISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|213868299|gb|ACJ54196.1|         DHKKTISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
gi|213868301|gb|ACJ54197.1|         DHKKTISSAIINHAFLQNTVMKNCNYKRKRRERDWDCNTKKDVCIPDRRY
                                                                                    :***

gi|11120233|gb|AAG30847.1|          QLCMKELTNLVNNTDTNFHSDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|12407747|gb|AAG53634.1|          QLCMKELTNLVNNTDTNFHSDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|18766659|gb|AAL79050.1|AF46      QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|18766819|gb|AAL79130.1|AF46      QLCMKELTNLVNNTDTNFHSDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|63146462|gb|AAY34059.1|          QLCMKELTNLVNNTDTNFHSDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|63146616|gb|AAY34136.1|          QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|11138983|gb|AAG31571.1|          QLCMKELTNLVNNTDTNFHSDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|73697858|gb|AAZ81532.1|          QLCMKELTNLVNNTDTNFHSDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|73697846|gb|AAZ81526.1|          QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|12004297|gb|AAG43989.1|          QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|12004299|gb|AAG43990.1|          QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|73697866|gb|AAZ81536.1|          QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|73697848|gb|AAZ81527.1|          QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|73697856|gb|AAZ81531.1|          QLCMKELTNLVNNTDTNFHSDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|73697862|gb|AAZ81534.1|          QLCMKELTNLVNNTDTNLHSDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|160276|gb|AAA63423.1|            QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|73697844|gb|AAZ81525.1|          QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|339961368|pdb|3RRC|A             QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|339961369|pdb|3RRC|B             QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|255914751|dbj|BAH96577.1|        QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926624|gb|ACN69874.1|         QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926626|gb|ACN69875.1|         QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926628|gb|ACN69876.1|         QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926630|gb|ACN69877.1|         QLCMKELTNLVNNTDTNFHSDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926632|gb|ACN69878.1|         QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926634|gb|ACN69879.1|         QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926636|gb|ACN69880.1|         QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926638|gb|ACN69881.1|         QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926640|gb|ACN69882.1|         QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926642|gb|ACN69883.1|         QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926644|gb|ACN69884.1|         QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926646|gb|ACN69885.1|         QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926648|gb|ACN69886.1|         QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926650|gb|ACN69887.1|         QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926652|gb|ACN69888.1|         QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926654|gb|ACN69889.1|         QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926656|gb|ACN69890.1|         QLCMKELTNLVNNTDPNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926658|gb|ACN69891.1|         QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926660|gb|ACN69892.1|         QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926662|gb|ACN69893.1|         QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926664|gb|ACN69894.1|         QLCMKELTNLVNNTDTNLHSDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926666|gb|ACN69895.1|         QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926668|gb|ACN69896.1|         QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926670|gb|ACN69897.1|         QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926672|gb|ACN69898.1|         QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926674|gb|ACN69899.1|         QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
```

782

gi|224926676|gb|ACN69900.1| QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926678|gb|ACN69901.1| QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926680|gb|ACN69902.1| QLCMKELTNLVNNTDTNLHSDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926682|gb|ACN69903.1| QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926684|gb|ACN69904.1| QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926686|gb|ACN69905.1| QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926688|gb|ACN69906.1| QLCMKELTNLVNNTDTNFHSDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926690|gb|ACN69907.1| QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926692|gb|ACN69908.1| QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926694|gb|ACN69909.1| QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926696|gb|ACN69910.1| QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926698|gb|ACN69911.1| QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926700|gb|ACN69912.1| QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926702|gb|ACN69913.1| QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926704|gb|ACN69914.1| QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926706|gb|ACN69915.1| QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAIKGDLLLKLNNY
gi|224926708|gb|ACN69916.1| QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926710|gb|ACN69917.1| QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926714|gb|ACN69918.1| QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926716|gb|ACN69919.1| QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926718|gb|ACN69920.1| QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926720|gb|ACN69921.1| QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926722|gb|ACN69922.1| QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926724|gb|ACN69923.1| QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926726|gb|ACN69924.1| QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926728|gb|ACN69925.1| QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926730|gb|ACN69926.1| QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926732|gb|ACN69927.1| QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926734|gb|ACN69928.1| QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926736|gb|ACN69929.1| QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926738|gb|ACN69930.1| QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926740|gb|ACN69931.1| QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926742|gb|ACN69932.1| QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926744|gb|ACN69933.1| QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926746|gb|ACN69934.1| QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926748|gb|ACN69935.1| QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926750|gb|ACN69936.1| QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926752|gb|ACN69937.1| QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926754|gb|ACN69938.1| QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLFKLNNY
gi|224926756|gb|ACN69939.1| QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926758|gb|ACN69940.1| QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926760|gb|ACN69941.1| QLCMKELTNLVNNTDTNFHSDITFRKLYLKRKLIYDAAVEGDLLFKLNNY
gi|224926762|gb|ACN69942.1| QLCMKELTNLVNNTDTNFHSDITFRKLYLKRKLIYDAAVEGDLLFKLNNY
gi|224926764|gb|ACN69943.1| QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926766|gb|ACN69944.1| QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLFKLNNY
gi|224926768|gb|ACN69945.1| QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926770|gb|ACN69946.1| QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926772|gb|ACN69947.1| QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLFKLNNY
gi|224926774|gb|ACN69948.1| QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926776|gb|ACN69949.1| QLCMKELTNLVNNTDTNFHSDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926778|gb|ACN69950.1| QLCMKELTNLVNNTDTNFHSDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926780|gb|ACN69951.1| QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926782|gb|ACN69952.1| QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926784|gb|ACN69953.1| QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926786|gb|ACN69954.1| QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926788|gb|ACN69955.1| QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926790|gb|ACN69956.1| QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926792|gb|ACN69957.1| QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
783 gi|224926794|gb|ACN69958.1| QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY

```
gi|224926796|gb|ACN69959.1|         QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926798|gb|ACN69960.1|         QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926800|gb|ACN69961.1|         QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926802|gb|ACN69962.1|         QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926804|gb|ACN69963.1|         QLCMKELTNLVNNTDTNFHSDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926806|gb|ACN69964.1|         QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926808|gb|ACN69965.1|         QLCMKELTNLVNNTDTNLHSDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926810|gb|ACN69966.1|         QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926812|gb|ACN69967.1|         QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926814|gb|ACN69968.1|         QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926816|gb|ACN69969.1|         QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926818|gb|ACN69970.1|         QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|224926820|gb|ACN69971.1|         QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|213868281|gb|ACJ54187.1|         QLCMKELTNLVNNTDTNLHSDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|213868283|gb|ACJ54188.1|         QLCMKELTNLVNNTDTNFHSDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|213868285|gb|ACJ54189.1|         QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|213868287|gb|ACJ54190.1|         QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|213868289|gb|ACJ54191.1|         QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|213868291|gb|ACJ54192.1|         QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|213868293|gb|ACJ54193.1|         QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|213868295|gb|ACJ54194.1|         QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|213868297|gb|ACJ54195.1|         QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|213868299|gb|ACJ54196.1|         QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
gi|213868301|gb|ACJ54197.1|         QLCMKELTNLVNNTDTNFHRDITFRKLYLKRKLIYDAAVEGDLLLKLNNY
                                    ***************.*:* ******************::****:*****

gi|11120233|gb|AAG30847.1|          RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNSRSIFGTGEKAQ
gi|12407747|gb|AAG53634.1|          RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTGEKAQ
gi|18766659|gb|AAL79050.1|AF46      RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTDEKAQ
gi|18766819|gb|AAL79130.1|AF46      RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTGEKAQ
gi|63146462|gb|AAY34059.1|          RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTGEKAQ
gi|63146616|gb|AAY34136.1|          RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTDEKAQ
gi|11138983|gb|AAG31571.1|          RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTGEKAQ
gi|73697858|gb|AAZ81532.1|          RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTDEKAQ
gi|73697846|gb|AAZ81526.1|          RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTGEKAQ
gi|12004297|gb|AAG43989.1|          RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSEVVENNLRSIFGTGEKAQ
gi|12004299|gb|AAG43990.1|          RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTDEKAQ
gi|73697866|gb|AAZ81536.1|          RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTGEKAQ
gi|73697848|gb|AAZ81527.1|          RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTDEKAQ
gi|73697856|gb|AAZ81531.1|          RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTGKNAQ
gi|73697862|gb|AAZ81534.1|          RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTGKNAQ
gi|160276|gb|AAA63423.1|            RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTDEKAQ
gi|73697844|gb|AAZ81525.1|          RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTDEKAQ
gi|339961368|pdb|3RRC|A             RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTDEKAQ
gi|339961369|pdb|3RRC|B             RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTDEKAQ
gi|255914751|dbj|BAH96577.1|        RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTGEKAQ
gi|224926624|gb|ACN69874.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSEVVENNLRSIFGTGEQAQ
gi|224926626|gb|ACN69875.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSEVVENNLRSIFGTGEKAQ
gi|224926628|gb|ACN69876.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTGEKAQ
gi|224926630|gb|ACN69877.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTGKNAQ
gi|224926632|gb|ACN69878.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTGEKAQ
gi|224926634|gb|ACN69879.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSEVVENNLRSIFGTGKNAQ
gi|224926636|gb|ACN69880.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSEVVENNLRSIFGTGEKAQ
gi|224926638|gb|ACN69881.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTGEKAQ
gi|224926640|gb|ACN69882.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSEVVENNLRSIFGTGKNAQ
gi|224926642|gb|ACN69883.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSEVVENNLRSIFGTGEKAQ
gi|224926644|gb|ACN69884.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSEVVENNLRSIFGTGEKAQ
gi|224926646|gb|ACN69885.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSEVVENNLRSIFGTGEKAQ
gi|224926648|gb|ACN69886.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTGEKAQ
```

784

```
gi|224926650|gb|ACN69887.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSEVVENNLRSIFGTGKNAQ
gi|224926652|gb|ACN69888.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSEVVENNLRSIFGTGEKAQ
gi|224926654|gb|ACN69889.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSEVVENNLRSIFGTGEKAQ
gi|224926656|gb|ACN69890.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTGEKAQ
gi|224926658|gb|ACN69891.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTGEKAQ
gi|224926660|gb|ACN69892.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSEVVENNLRSIFGTGKNAQ
gi|224926662|gb|ACN69893.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTGEKAQ
gi|224926664|gb|ACN69894.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTGKNAQ
gi|224926666|gb|ACN69895.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTGEKAQ
gi|224926668|gb|ACN69896.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSEVVENNLRSIFGTGKNAQ
gi|224926670|gb|ACN69897.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTGEKAQ
gi|224926672|gb|ACN69898.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTGEKAQ
gi|224926674|gb|ACN69899.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSEVVENNLRSIFGTGKNAQ
gi|224926676|gb|ACN69900.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSEVVENNLRSIFGTGKNAQ
gi|224926678|gb|ACN69901.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTGEKAQ
gi|224926680|gb|ACN69902.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTGKNAQ
gi|224926682|gb|ACN69903.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTGEKAQ
gi|224926684|gb|ACN69904.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTDEKAQ
gi|224926686|gb|ACN69905.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTDEKAQ
gi|224926688|gb|ACN69906.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTGKNAQ
gi|224926690|gb|ACN69907.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSEVVENNLRSIFGTGKNAQ
gi|224926692|gb|ACN69908.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTDEKAQ
gi|224926694|gb|ACN69909.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTGKNAQ
gi|224926696|gb|ACN69910.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSEVVENNLRSIFGTGKNAQ
gi|224926698|gb|ACN69911.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTGEKAQ
gi|224926700|gb|ACN69912.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTGEKAQ
gi|224926702|gb|ACN69913.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSEVVENNLRSIFGTGKNAQ
gi|224926704|gb|ACN69914.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTGEKAQ
gi|224926706|gb|ACN69915.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSEVVENNLRSIFGTGKNAQ
gi|224926708|gb|ACN69916.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSEVVENNLRSIFGTGKNAQ
gi|224926710|gb|ACN69917.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTDEKAQ
gi|224926714|gb|ACN69918.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTDEKAQ
gi|224926716|gb|ACN69919.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTGEKAQ
gi|224926718|gb|ACN69920.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTDEKAQ
gi|224926720|gb|ACN69921.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTDEKAQ
gi|224926722|gb|ACN69922.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTGEKAQ
gi|224926724|gb|ACN69923.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTGEKAQ
gi|224926726|gb|ACN69924.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSEVVENNLRSIFGTGKNAQ
gi|224926728|gb|ACN69925.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTGEKAQ
gi|224926730|gb|ACN69926.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTDEKAQ
gi|224926732|gb|ACN69927.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSEVVENNLRSIFGTGKNAQ
gi|224926734|gb|ACN69928.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTGEKAQ
gi|224926736|gb|ACN69929.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTGEKAQ
gi|224926738|gb|ACN69930.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSEVVENNLRSIFGTGEQAQ
gi|224926740|gb|ACN69931.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSEVVENNLRSIFGTGEQAQ
gi|224926742|gb|ACN69932.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTDEKAQ
gi|224926744|gb|ACN69933.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSEVVENNLRSIFGTGEKAQ
gi|224926746|gb|ACN69934.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTGEKAQ
gi|224926748|gb|ACN69935.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTGEKAQ
gi|224926750|gb|ACN69936.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTGEKAQ
gi|224926752|gb|ACN69937.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTGEKAQ
gi|224926754|gb|ACN69938.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVEDNLRSIFGTGKNAQ
gi|224926756|gb|ACN69939.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTGEKAQ
gi|224926758|gb|ACN69940.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSEVVENNLRSIFGTGEQAQ
gi|224926760|gb|ACN69941.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVEDNLRSIFGTGKNAQ
gi|224926762|gb|ACN69942.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVEDNLRSIFGTGKNAQ
gi|224926764|gb|ACN69943.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTGEKAQ
gi|224926766|gb|ACN69944.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVEDNLRSIFGTGKNAQ
gi|224926768|gb|ACN69945.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTDEKAQ
```

```
gi|224926770|gb|ACN69946.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSEVVENNLRSIFGTGEQAQ
gi|224926772|gb|ACN69947.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSEVVENNLRSIFGTGKNAQ
gi|224926774|gb|ACN69948.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTDEKAQ
gi|224926776|gb|ACN69949.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTGKNAQ
gi|224926778|gb|ACN69950.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTGKNAQ
gi|224926780|gb|ACN69951.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTGEKAQ
gi|224926782|gb|ACN69952.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTGEKAQ
gi|224926784|gb|ACN69953.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIRYSEVVENILRSIFGTGEQAQ
gi|224926786|gb|ACN69954.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTGEKAQ
gi|224926788|gb|ACN69955.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTGEKAQ
gi|224926790|gb|ACN69956.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTDEKAQ
gi|224926792|gb|ACN69957.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTGEKAQ
gi|224926794|gb|ACN69958.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTGEKAQ
gi|224926796|gb|ACN69959.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTGEKAQ
gi|224926798|gb|ACN69960.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTGEKAQ
gi|224926800|gb|ACN69961.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTDEKAQ
gi|224926802|gb|ACN69962.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSEVVENNLRSIFGTGKNAQ
gi|224926804|gb|ACN69963.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSEVVENNLRSIFGTGKNAQ
gi|224926806|gb|ACN69964.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTGEKAQ
gi|224926808|gb|ACN69965.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTGKNAQ
gi|224926810|gb|ACN69966.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSEVVENNLRSIFGTGKNAQ
gi|224926812|gb|ACN69967.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSEVVENNLRSIFGTGKNAQ
gi|224926814|gb|ACN69968.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTGEKAQ
gi|224926816|gb|ACN69969.1|         KYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTGEKAQ
gi|224926818|gb|ACN69970.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTGEKAQ
gi|224926820|gb|ACN69971.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTGEKAQ
gi|213868281|gb|ACJ54187.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTGKNAQ
gi|213868283|gb|ACJ54188.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTGKNAQ
gi|213868285|gb|ACJ54189.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSEVVENNLRSIFGTGEKAQ
gi|213868287|gb|ACJ54190.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSEVVENNLRSIFGTGEKAQ
gi|213868289|gb|ACJ54191.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTDEKAQ
gi|213868291|gb|ACJ54192.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTGEKAQ
gi|213868293|gb|ACJ54193.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTDEKAQ
gi|213868295|gb|ACJ54194.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTDEKAQ
gi|213868297|gb|ACJ54195.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTDEKAQ
gi|213868299|gb|ACJ54196.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTDEKAQ
gi|213868301|gb|ACJ54197.1|         RYNKDFCKDIRWSLGDFGDIIMGTDMEGIGYSKVVENNLRSIFGTGEKAQ
                                    :*************************** **:***:  ******.::**

gi|11120233|gb|AAG30847.1|          QHRKQRWNESKAQIWTAMMYSVKKRLKGNFIWICKINVAVNIEPQIYRWI
gi|12407747|gb|AAG53634.1|          QHRKQWWNESKAQIWTAMMYSVKKRLKGNFIWICKINVAVNIEPQIYRWI
gi|18766659|gb|AAL79050.1|AF46      QRRKQWWNESKAQIWTAMMYSVKKRLKGNFIWICKLNVAVNIEPQIYRWI
gi|18766819|gb|AAL79130.1|AF46      QHRKQWWNESKAQIWTAMMYSVKKRLKGNFIWICKINVAVNIEPQIYRWI
gi|63146462|gb|AAY34059.1|          QHRKQWWNESKAQIWTAMMYSVKKRLKGNFIWICKINVAVNIEPQIYRWI
gi|63146616|gb|AAY34136.1|          QRRKQWWNESKAQIWTAMMYSVKKRLKGNFIWICKLNVAVNIEPQIYRWI
gi|11138983|gb|AAG31571.1|          QHRKQRWNESKAQIWTAMMYSVKKRLKGNFIWICKINVAVNIEPQIYRWI
gi|73697858|gb|AAZ81532.1|          QRRKQWWNESKAQIWTAMMYSVKKRLKGKFIWICKINVAVNIEPQIYRRI
gi|73697846|gb|AAZ81526.1|          QHRKQWWNESKAQIWTAMMYSVKKRLKGNFIWICKLNVAVNIEPQIYRWI
gi|12004297|gb|AAG43989.1|          QHRKQWWNESKAQIWTAMMYSVKKRLKGKFIWICKINVAVNIEPQIYRRI
gi|12004299|gb|AAG43990.1|          QRRKQWWNESKAQIWTAMMYSVKKRLKGKFIWICKINVAVNIEPQIYRRI
gi|73697866|gb|AAZ81536.1|          QRHKQWWNESKAQIWTAMMYSVKKRLKGKFIWICKINVAVNIEPQIYRRI
gi|73697848|gb|AAZ81527.1|          QRRKQWWNESKAQIWTAMMYSVKKRLKGNFIWICKLNVAVNIEPQIYRWI
gi|73697856|gb|AAZ81531.1|          QHRKQWWNESKAQIWTAMMYSVKKRLKGKFIWICKINVAVNIEPQIYRRI
gi|73697862|gb|AAZ81534.1|          QHRKQWWNETKAQIWRAMMYSVKKRLKGNFIWICKINVAVNIEPQIYRWI
gi|160276|gb|AAA63423.1|            QRRKQWWNESKAQIWTAMMYSVKKRLKGNFIWICKLNVAVNIEPQIYRWI
gi|73697844|gb|AAZ81525.1|          QRRKQWWNESKAQIWTAMMYSVKKRLKGNFIWICKLNVAVNIEPQIYRWI
gi|339961368|pdb|3RRC|A             QRRKQWWNESKAQIWTAMMYSVKKRLKGNFIWICKLNVAVNIEPQIYRWI
gi|339961369|pdb|3RRC|B             QRRKQWWNESKAQIWTAMMYSVKKRLKGNFIWICKLNVAVNIEPQIYRWI
gi|255914751|dbj|BAH96577.1|        QHRKQWWNESKAQIWTAMMYSVKKRLKGNFIWICKINVAVNIEPQIYRWI
```

786

```
gi|224926624|gb|ACN69874.1|         QRRKQWWNESKAQIWTAMMYSVKKRLKGKFIWICKINVAVNIEPQIYRRI
gi|224926626|gb|ACN69875.1|         QRRKQWWNESKAQIWTAMMYSVKKRLKGKFMWICKINVAVNIEPQIYRRI
gi|224926628|gb|ACN69876.1|         QHRKQWWNESKAQIWTAMMYSVKKRLKGNFIWICKLNVAVNIEPQIYRWI
gi|224926630|gb|ACN69877.1|         QHRKQWWNETKAQIWRAMMYSVKKRLKGNFIWICKINVAVNIEPQIYRWI
gi|224926632|gb|ACN69878.1|         QHRKQWWNESKAQIWTAMMYSVKKRLKGNFIWICKLNVAVNIEPQIYRWI
gi|224926634|gb|ACN69879.1|         QRRKQWWNESKAQIWTAMMYSVKKRLKGKFIWICKINVAVNIEPQIYRRI
gi|224926636|gb|ACN69880.1|         QRRKQWWNESKAQIWTAMMYSVKKRLKGKFMWICKINVAVNIEPQIYRRI
gi|224926638|gb|ACN69881.1|         QHRKQWWNESKAQIWTAMMYSVKKRLKGNFIWICKLNVAVNIEPQIYRWI
gi|224926640|gb|ACN69882.1|         QHRKQWWNESKAQIWTAMMYSVKKRLKGKFIWICKINVAVNIEPQIYRRI
gi|224926642|gb|ACN69883.1|         QRRKQWWNESKAQIWTAMMYSVKKRLKGKFMWICKINVAVNIEPQIYRRI
gi|224926644|gb|ACN69884.1|         QRRKQWWNESKAQIWTAMMYSVKKRLKGKFMWICKINVAVNIEPQIYRRI
gi|224926646|gb|ACN69885.1|         QRRKQWWNESKAQIWTAMMYSVKKRLKGKFIWICKINVAVNIEPQIYRRI
gi|224926648|gb|ACN69886.1|         QHRKQWWNESKAQIWTAMMYSVKKRLKGNFIWICKLNVAVNIEPQIYRWI
gi|224926650|gb|ACN69887.1|         QRRKQWWNESKAQIWTAMMYSVKKRLKGKFIWICKINVAVNIEPQIYRRI
gi|224926652|gb|ACN69888.1|         QRRKQWWNESKAQIWTAMMYSVKKRLKGKFIWICKINVAVNIEPQIYRRI
gi|224926654|gb|ACN69889.1|         QRRKQWWNESKAQIWTAMMYSVKKRLKGKFMWICKINVAVNIEPQIYRRI
gi|224926656|gb|ACN69890.1|         QHRKQWWNESKAQIWTAMMYSVKKRLKGNFIWICKLNVAVNIEPQIYRWI
gi|224926658|gb|ACN69891.1|         QHRKQWWNESKAQIWTAMMYSVKKRLKGNFIWICKLNVAVNIEPQIYRWI
gi|224926660|gb|ACN69892.1|         QHRKQWWNESKAQIWTAMMYSVKKRLKGKFIWICKINVAVNIEPQIYRRI
gi|224926662|gb|ACN69893.1|         QHRKQWWNESKAQIWTAMMYSVKKRLKGNFIWICKLNVAVNIEPQIYRWI
gi|224926664|gb|ACN69894.1|         QHRKQWWNETKAQIWRAMMYSVKKRLKGNFIWICKINVAVNIEPQIYRWI
gi|224926666|gb|ACN69895.1|         QHRKQWWNESKAQIWTAMMYSVKKRLKGNFIWICKLNVAVNIEPQIYRWI
gi|224926668|gb|ACN69896.1|         QRRKQWWNESKAQIWTAMMYSVKKRLKGKFIWICKINVAVNIEPQIYRRI
gi|224926670|gb|ACN69897.1|         QHRKQWWNESKAQIWTAMMYSVKKRLKGNFIWICKLNVAVNIEPQIYRWI
gi|224926672|gb|ACN69898.1|         QHRKQWWNESKAQIWTAMMYSVKKRLKGNFIWICKLNVAVNIEPQIYRWI
gi|224926674|gb|ACN69899.1|         QHRKQWWNESKAQIWTAMMYSVKKRLKGKFIWICKINVAVNIEPQIYRRI
gi|224926676|gb|ACN69900.1|         QHRKQWWNESKAQIWTAMMYSVKKRLKGKFIWICKINVAVNIEPQIYRRI
gi|224926678|gb|ACN69901.1|         QHRKQWWNESKAQIWTAMMYSVKKRLKGNFIWICKLNVAVNIEPQIYRWI
gi|224926680|gb|ACN69902.1|         QHRKQWWNETKAQIWRAMMYSVKKRLKGNFIWICKINVAVNIEPQIYRWI
gi|224926682|gb|ACN69903.1|         QHRKQWWNESKAQIWTAMMYSVKKRLKGNFIWICKLNVAVNIEPQIYRWI
gi|224926684|gb|ACN69904.1|         QRCKQWWNESKAQIWTAMMYSVKKRLKGNFIWICKLNVAVNIEPQIYRWI
gi|224926686|gb|ACN69905.1|         QRRKQWWNESKAQIWTAMMYSVKKRLKGNFIWICKLNVAVNIEPQIYRWI
gi|224926688|gb|ACN69906.1|         QHRKQWWNETKAQIWRAMMYSVKKRLKGNFIWICKINVAVNIEPQIYRWI
gi|224926690|gb|ACN69907.1|         QHRKQWWNESKAQIWTAMMYSVKKRLKGKFIWICKINVAVNIEPQIYRRI
gi|224926692|gb|ACN69908.1|         QRRKQWWNESKAQIWTAMMYSVKKRLKGNFIWICKLNVAVNIEPQIYRWI
gi|224926694|gb|ACN69909.1|         QRCKQWWNEFKAQIWTAMMYSVKKRLKGKFIWICKINVAVNIEPQIYRRI
gi|224926696|gb|ACN69910.1|         QRCKQWWNESKAQIWTAMMYSVKKRLKGKFIWICKINVAVNIEPQIYRRI
gi|224926698|gb|ACN69911.1|         QHCKQWWNESKAQIWTAMMYSVKKRLKGNFIWICKLNVAVNIEPQIYRWI
gi|224926700|gb|ACN69912.1|         QHCKQWWNESKAQIWTAMMYSVKKRLKGNFIWICKLNVAVNIEPQIYRWI
gi|224926702|gb|ACN69913.1|         QHRKQWWNESKAQIWTAMMYSVKKRLKGKFIWICKINVAVNIEPQIYRRI
gi|224926704|gb|ACN69914.1|         QHRKQWWNESKAQIWTAMMYSVKKRLKGNFIWICKLNVAVNIEPQIYRWI
gi|224926706|gb|ACN69915.1|         QHRKQWWNESKAQIWTAMMYSVKKRLKGKFMWICKINVAVNIEPQIYRRI
gi|224926708|gb|ACN69916.1|         QHRKQWWNESKAQIWTAMMYSVKKRLKGKFIWICKINVAVNIEPQIYRRI
gi|224926710|gb|ACN69917.1|         QRRKQWWNESKAQIWTAMMYSVKKRLKGNFIWICKLNVAVNIEPQIYRWI
gi|224926714|gb|ACN69918.1|         QRRKQWWNESKAQIWTAMMYSVKKRLKGNFIWICKLNVAVNIEPQIYRWI
gi|224926716|gb|ACN69919.1|         QHRKQWWNESKAQIWTAMMYSVKKRLKGNFIWICKLNVAVNIEPQIYRWI
gi|224926718|gb|ACN69920.1|         QRRKQWWNESKAQIWTAMMYSVKKRLKGNFIWICKLNVAVNIEPQIYRWI
gi|224926720|gb|ACN69921.1|         QRRKQWWNESKAQIWTAMMYSVKKRLKGNFIWICKLNVAVNIEPQIYRWI
gi|224926722|gb|ACN69922.1|         QHRKQWWNESKAQIWTAMMYSVKKRLKGNFIWICKLNVAVNIEPQIYRWI
gi|224926724|gb|ACN69923.1|         QHRKQWWNESKAQIWTAMMYSVKKRLKGNFIWICKLNVAVNIEPQIYRWI
gi|224926726|gb|ACN69924.1|         QHRKQWWNESKAQIWTAMMYSVKKRLKGKFIWICKINVAVNIEPQIYRRI
gi|224926728|gb|ACN69925.1|         QHRKQWWNESKAQIWTAMMYSVKKRLKGNFIWICKLNVAVNIEPQIYRWI
gi|224926730|gb|ACN69926.1|         QRRKQWWNESKAQIWTAMMYSVKKRLKGNFIWICKLNVAVNIEPQIYRWI
gi|224926732|gb|ACN69927.1|         QHRKQWWNESKAQIWTAMMYSVKKRLKGKFIWICKINVAVNIEPQIYRRI
gi|224926734|gb|ACN69928.1|         QHRKQWWNESKAQIWTAMMYSVKKRLKGNFIWICKLNVAVNIEPQIYRWI
gi|224926736|gb|ACN69929.1|         QHRKQWWNESKAQIWTAMMYSVKKRLKGNFIWICKINVAVNIEPQIYRWI
gi|224926738|gb|ACN69930.1|         QRRKQWWNESKAQIWTAMMYSVKKRLKGKFIWICKINVAVNIEPQIYRRI
gi|224926740|gb|ACN69931.1|         QRRKQWWNESKAQIWTAMMYSVKKRLKGKFIWICKLNVAVNIEPQIYRRI
787 gi|224926742|gb|ACN69932.1|         QRRKQWWNESKAQIWTAMMYSVKKRLKGNFIWICKLNVAVNTEPQIYRWI
```

```
gi|224926744|gb|ACN69933.1|          QRRKQWWNESKAQIWTAMMYSVKKRLKGKFMWICKINVAVNIEPQIYRRI
gi|224926746|gb|ACN69934.1|          QHRKQWWNESKAQIWTAMMYSVKKRLKGNFIWICKLNVAVNIEPQIYRWI
gi|224926748|gb|ACN69935.1|          QHRKQWWNESKAQIWTAMMYSVKKRLKGNFIWICKLNVAVNIEPQIYRWI
gi|224926750|gb|ACN69936.1|          QHRKQWWNESKAQIWTAMMYSVKKRLKGNFIWICKLNVAVNIEPQIYRWI
gi|224926752|gb|ACN69937.1|          QHRKQWWNESKAQIWTAMMYSVKKRLKGNFIWICKLNVAVNIEPQIYRWI
gi|224926754|gb|ACN69938.1|          QHRKQWWNESKAQIWTAMMYSVKKRLKGKFIWICKINVAVNIEPQIYRRI
gi|224926756|gb|ACN69939.1|          QHRKQWWNESKAQIWTAMMYSVKKRLKGNFIWICKLNVAVNIEPQIYRWI
gi|224926758|gb|ACN69940.1|          QRCKQWWNEFKAQIWTAMMYSVKKRLKGKFIWICKINVAVNIEPQIYRRI
gi|224926760|gb|ACN69941.1|          QHRKQWWNESKAQIWTAMMYSVKKRLKGKFIWICKINVAVNIEPQIYRRI
gi|224926762|gb|ACN69942.1|          QHRKQWWNESKAQIWTAMMYSVKKRLKGKFIWICKINVAVNIEPQIYRRI
gi|224926764|gb|ACN69943.1|          QHRKQWWNESKAQIWTAMMYSVKKRLKGNFIWICKLNVAVNIEPQIYRWI
gi|224926766|gb|ACN69944.1|          QHRKQWWNESKAQIWTAMMYSVKKRLKGKFIWICKINVAVNIEPQIYRRI
gi|224926768|gb|ACN69945.1|          QRRKQWWNESKAQIWTAMMYSVKKRLKGNFIWICKLNVAVNIEPQIYRWI
gi|224926770|gb|ACN69946.1|          QRRKQWWNESKAQIWTAMMYSVKKRLKGKFIWICKINVAVNIEPQIYRRI
gi|224926772|gb|ACN69947.1|          QRRKQWWNESKAQIWTAMMYSVKKRLKGKFIWICKINVAVNIEPQIYRRI
gi|224926774|gb|ACN69948.1|          QRRKQWWNESKAQIWTAMMYSVKKRLKGNFIWICKLNVAVNIEPQIYRWI
gi|224926776|gb|ACN69949.1|          QHRKQWWNETKAQIWRAMMYSVKKRLKGNFIWICKINVAVNIEPQIYRWI
gi|224926778|gb|ACN69950.1|          QHRKQWWNETKAQIWRAMMYSVKKRLKGNFIWICKINVAVNIEPQIYRWI
gi|224926780|gb|ACN69951.1|          QHRKQWWNESKAQIWTAMMYSVKKRLKGNFIWICKLNVAVNIEPQIYRWI
gi|224926782|gb|ACN69952.1|          QHRKQWWNESKAQIWTAMMYSVKKRLKGNFIWICKLNVAVNIEPQIYRWI
gi|224926784|gb|ACN69953.1|          QRRKQWWNESKAQIWTAMMYSVKKRLKGKFIWICKINVAVNIDPQIYRRF
gi|224926786|gb|ACN69954.1|          QHRKQWWNESKAQIWTAMMYSVKKRLKGNFIWICKLNVAVNIEPQIYRWI
gi|224926788|gb|ACN69955.1|          QHRKQWWNESKAQIWTAMMYSVKKRLKGNFIWICKLNVAVNIEPQIYRWI
gi|224926790|gb|ACN69956.1|          QRRKQWWNESKAQIWTAMMYSVKKRLKGNFIWICKLNVAVNIEPQIYRWI
gi|224926792|gb|ACN69957.1|          QHRKQWWNESKAQIWTAMMYSVKKRLKGNFIWICKLNVAVNIEPQIYRWI
gi|224926794|gb|ACN69958.1|          QHRKQWWNESKAQIWTAMMYSVKKRLKGNFIWICKLNVAVNIEPQIYRWI
gi|224926796|gb|ACN69959.1|          QHRKQWWNESKAQIWTAMMYSVKKRLKGNFIWICKLNVAVNIEPQIYRWI
gi|224926798|gb|ACN69960.1|          QHRKQWWNESKAQIWTAMMYSVKKRLKGNFIWICKLNVAVNIEPQIYRWI
gi|224926800|gb|ACN69961.1|          QRRKQWWNESKAQIWTAMMYSVKKRLKGNFIWICKLNVAVNIEPQIYRWI
gi|224926802|gb|ACN69962.1|          QHRKQWWNESKAQIWTAMMYSVKKRLKGKFIWICKINVAVNIEPQIYRRI
gi|224926804|gb|ACN69963.1|          QRRKQWWNESKAQIWTAMMYSVKKRLKGKFIWICKINVAVNIEPQIYRRI
gi|224926806|gb|ACN69964.1|          QHRKQWWNESKAQIWTAMMYSVKKRLKGNFIWICKLNVAVNIEPQIYRWI
gi|224926808|gb|ACN69965.1|          QHRKQWWNETKAQIWRAMMYSVKKRLKGNFIWICKINVAVNIEPQIYRWI
gi|224926810|gb|ACN69966.1|          QRRKQWWNESKAQIWTAMMYSVKKRLKGKFIWICKINVAVNIEPQIYRRI
gi|224926812|gb|ACN69967.1|          QHRKQWWNESKAQIWTAMMYSVKKRLKGKFIWICKINVAVNIEPQIYRRI
gi|224926814|gb|ACN69968.1|          QHRKQWWNESKAQIWTAMMYSVKKRLKGNFIWICKLNVAVNIEPQIYRWI
gi|224926816|gb|ACN69969.1|          QHRKQWWNESKAQIWTAMMYSVKKRLKGNFIWICKLNVAVNIEPQIYRWI
gi|224926818|gb|ACN69970.1|          QHRKQWWNESKAQIWTAMMYSVKKRLKGNFIWICKLNVAVNIEPQIYRWI
gi|224926820|gb|ACN69971.1|          QHRKQWWNESKAQIWTAMMYSVKKRLKGNFIWICKLNVAVNIEPQIYRWI
gi|213868281|gb|ACJ54187.1|          QHRKQWWNETKAQIWRAMMYSVKKRLKGNFIWICKINVAVNIEPQIYRWI
gi|213868283|gb|ACJ54188.1|          QHRKQWWNETKAQIWRAMMYSVKKRLKGNFIWICKINVAVNIEPQIYRWI
gi|213868285|gb|ACJ54189.1|          QRRKQWWNESKAQIWTAMMYSVKKRLKGKFMWICKINVAVNIEPQIYRRI
gi|213868287|gb|ACJ54190.1|          QRRKQWWNESKAQIWTAMMYSVKKRLKGKFIWICKINVAVNIEPQIYRRI
gi|213868289|gb|ACJ54191.1|          QRRKQWWNESKAQIWTAMMYSVKKRLKGKFIWICKINVAVNIEPQIYRRI
gi|213868291|gb|ACJ54192.1|          QHRKQWWNESKAQIWTAMMYSVKKRLKGNFIWICKLNVAVNIEPQIYRWI
gi|213868293|gb|ACJ54193.1|          QRRKQWWNESKAQIWTAMMYSVKKRLKGNFIWICKLNVAVNIEPQIYRWI
gi|213868295|gb|ACJ54194.1|          QRRKQWWNESKAQIWTAMMYSVKKRLKGNFIWICKLNVAVNIEPQIYRRI
gi|213868297|gb|ACJ54195.1|          QRRKQWWNESKAQIWTAMMYSVKKRLKGNFIWICKLNVAVNIEPQIYRWI
gi|213868299|gb|ACJ54196.1|          QRRKQWWNESKAQIWTAMMYSVKKRLKGKFIWICKINVAVNIEPQIYRRI
gi|213868301|gb|ACJ54197.1|          QHRKQWWNESKAQIWTAMMYSVKKRLKGNFIWICKLNVAVNIEPQIYRWI
                                     *: ** *** ***** ************:*:****:***** :***** :

gi|11120233|gb|AAG30847.1|           REWGRDYVKELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|12407747|gb|AAG53634.1|           REWGRDYVKELPTEVQKLKEKCDGKINYPDKKVCKV-PPCQNACKSYDQW
gi|18766659|gb|AAL79050.1|AF46       REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|18766819|gb|AAL79130.1|AF46       REWGRDYVKELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|63146462|gb|AAY34059.1|           REWGRDYVKELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|63146616|gb|AAY34136.1|           REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|11138983|gb|AAG31571.1|           REWGRDYVKELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
```

```
gi|73697858|gb|AAZ81532.1|          REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|73697846|gb|AAZ81526.1|          REWGRDYVSELPTEVQKLKEKCDGKFNYTDKKVCKV-PPCQNACKSYDQW
gi|12004297|gb|AAG43989.1|          REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|12004299|gb|AAG43990.1|          REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|73697866|gb|AAZ81536.1|          REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|73697848|gb|AAZ81527.1|          REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|73697856|gb|AAZ81531.1|          REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|73697862|gb|AAZ81534.1|          REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKVLPPCQNACKSYDQW
gi|160276|gb|AAA63423.1|            REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|73697844|gb|AAZ81525.1|          REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|339961368|pdb|3RRC|A             REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|339961369|pdb|3RRC|B             REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|255914751|dbj|BAH96577.1|        REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926624|gb|ACN69874.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926626|gb|ACN69875.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926628|gb|ACN69876.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926630|gb|ACN69877.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKVLPPCQNACKSYDQW
gi|224926632|gb|ACN69878.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926634|gb|ACN69879.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926636|gb|ACN69880.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926638|gb|ACN69881.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926640|gb|ACN69882.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926642|gb|ACN69883.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926644|gb|ACN69884.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926646|gb|ACN69885.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926648|gb|ACN69886.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926650|gb|ACN69887.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926652|gb|ACN69888.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926654|gb|ACN69889.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926656|gb|ACN69890.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926658|gb|ACN69891.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926660|gb|ACN69892.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926662|gb|ACN69893.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926664|gb|ACN69894.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKVLPPCQNACKSYDQW
gi|224926666|gb|ACN69895.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926668|gb|ACN69896.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926670|gb|ACN69897.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926672|gb|ACN69898.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926674|gb|ACN69899.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926676|gb|ACN69900.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926678|gb|ACN69901.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926680|gb|ACN69902.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKVLPPCQNACKSYDQW
gi|224926682|gb|ACN69903.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926684|gb|ACN69904.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926686|gb|ACN69905.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926688|gb|ACN69906.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKVLPPCQNACKSYDQW
gi|224926690|gb|ACN69907.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926692|gb|ACN69908.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926694|gb|ACN69909.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926696|gb|ACN69910.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926698|gb|ACN69911.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926700|gb|ACN69912.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926702|gb|ACN69913.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926704|gb|ACN69914.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926706|gb|ACN69915.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926708|gb|ACN69916.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926710|gb|ACN69917.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926714|gb|ACN69918.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926716|gb|ACN69919.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
```

```
gi|224926718|gb|ACN69920.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926720|gb|ACN69921.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926722|gb|ACN69922.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926724|gb|ACN69923.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926726|gb|ACN69924.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926728|gb|ACN69925.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926730|gb|ACN69926.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926732|gb|ACN69927.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926734|gb|ACN69928.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926736|gb|ACN69929.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926738|gb|ACN69930.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926740|gb|ACN69931.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926742|gb|ACN69932.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926744|gb|ACN69933.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926746|gb|ACN69934.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDEW
gi|224926748|gb|ACN69935.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926750|gb|ACN69936.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926752|gb|ACN69937.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926754|gb|ACN69938.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926756|gb|ACN69939.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDEW
gi|224926758|gb|ACN69940.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926760|gb|ACN69941.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926762|gb|ACN69942.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926764|gb|ACN69943.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926766|gb|ACN69944.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926768|gb|ACN69945.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926770|gb|ACN69946.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926772|gb|ACN69947.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926774|gb|ACN69948.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926776|gb|ACN69949.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKVLPPCQNACKSYDQW
gi|224926778|gb|ACN69950.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKVLPPCQNACKSYDEW
gi|224926780|gb|ACN69951.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926782|gb|ACN69952.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926784|gb|ACN69953.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926786|gb|ACN69954.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926788|gb|ACN69955.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926790|gb|ACN69956.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926792|gb|ACN69957.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926794|gb|ACN69958.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926796|gb|ACN69959.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926798|gb|ACN69960.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926800|gb|ACN69961.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926802|gb|ACN69962.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926804|gb|ACN69963.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926806|gb|ACN69964.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926808|gb|ACN69965.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKVLPPCQNACKSYDQW
gi|224926810|gb|ACN69966.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926812|gb|ACN69967.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926814|gb|ACN69968.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926816|gb|ACN69969.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926818|gb|ACN69970.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|224926820|gb|ACN69971.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|213868281|gb|ACJ54187.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKVLPPCQNACKSYDQW
gi|213868283|gb|ACJ54188.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKVLPPCQNACKSYDQW
gi|213868285|gb|ACJ54189.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|213868287|gb|ACJ54190.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|213868289|gb|ACJ54191.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|213868291|gb|ACJ54192.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDEW
gi|213868293|gb|ACJ54193.1|         REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
```

790

```
gi|213868295|gb|ACJ54194.1|          REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|213868297|gb|ACJ54195.1|          REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|213868299|gb|ACJ54196.1|          REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDQW
gi|213868301|gb|ACJ54197.1|          REWGRDYVSELPTEVQKLKEKCDGKINYTDKKVCKV-PPCQNACKSYDEW
                                     ********.****************:**.******* ***********:*

gi|11120233|gb|AAG30847.1|           ITRKKNQWDVLSNKFISVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|12407747|gb|AAG53634.1|           ITRKKNQWDVLSNKFISVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|18766659|gb|AAL79050.1|AF46       ITRKKNQWDVLSNKFISVKNAEKVQTAGIVTPYD----------------
gi|18766819|gb|AAL79130.1|AF46       ITRKKNQWDVLSNKFISVKNAEKVQTAGIVTPYD----------------
gi|63146462|gb|AAY34059.1|           ITRKKNQWDVLSNKFISVKNAEKVQTAGIVTPYD----------------
gi|63146616|gb|AAY34136.1|           ITRKKNQWDVLSNKFISVKNAEKVQTAGIVTP------------------
gi|11138983|gb|AAG31571.1|           ITRKKNQWDVLSNKFISVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|73697858|gb|AAZ81532.1|           ITRKKNQWDVLSNKFKSVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|73697846|gb|AAZ81526.1|           ITRKKNQWDVLSNKFISVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|12004297|gb|AAG43989.1|           ITRKKNQWDVLSNKFKSVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|12004299|gb|AAG43990.1|           ITRKKNQWDVLSNKFKSVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|73697866|gb|AAZ81536.1|           ITRKKNQWDVLSNKFKSVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|73697848|gb|AAZ81527.1|           ITRKKNQWDVLSNKFISVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|73697856|gb|AAZ81531.1|           ITRKKNQWDVLSNKFKSVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|73697862|gb|AAZ81534.1|           ITRKKNQWDVLSNKFISVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|160276|gb|AAA63423.1|             ITRKKNQWDVLSNKFISVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|73697844|gb|AAZ81525.1|           ITRKKNQWDVLSNKFISVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|339961368|pdb|3RRC|A              ITRKKNQWDVLSNKFISVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|339961369|pdb|3RRC|B              ITRKKNQWDVLSNKFISVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|255914751|dbj|BAH96577.1|         ITRKKNQWDVLSNKFKSVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926624|gb|ACN69874.1|          ITRKKNQWDVLSNKFKSVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926626|gb|ACN69875.1|          ITRKKNQWDVLSNKFISVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926628|gb|ACN69876.1|          ITRKKNQWDVLSNKFKSVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926630|gb|ACN69877.1|          ITRKKNQWDVLSNKFISVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926632|gb|ACN69878.1|          ITRKKNQWDVLSNKFKSVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926634|gb|ACN69879.1|          ITRKKNQWDVLSNKFKSVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926636|gb|ACN69880.1|          ITRKKNQWDVLSNKFISVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926638|gb|ACN69881.1|          ITRKKNQWDVLSNKFKSVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926640|gb|ACN69882.1|          ITRKKNQWDVLSNKFISVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926642|gb|ACN69883.1|          ITRKKNQWDVLSNKFISVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926644|gb|ACN69884.1|          ITRKKNQWDVLSNKFISVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926646|gb|ACN69885.1|          ITRKKNQWDVLSNKFKSVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926648|gb|ACN69886.1|          ITRKKNQWDVLSNKFKSVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926650|gb|ACN69887.1|          ITRKKNQWDVLSNKFKSVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926652|gb|ACN69888.1|          ITRKKNQWDVLSNKFKSVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926654|gb|ACN69889.1|          ITRKKNQWDVLSNKFISVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926656|gb|ACN69890.1|          ITRKKNQWDVLSNKFKSVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926658|gb|ACN69891.1|          ITRKKNQWDVLSNKFKSVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926660|gb|ACN69892.1|          ITRKKNQWDVLSNKFISVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926662|gb|ACN69893.1|          ITRKKNQWDVLSNKFKSVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926664|gb|ACN69894.1|          ITRKKNQWDVLSNKFISVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926666|gb|ACN69895.1|          ITRKKNQWDVLSNKFKSVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926668|gb|ACN69896.1|          ITRKKNQWDVLSNKFKSVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926670|gb|ACN69897.1|          ITRKKNQWDVLSNKFKSVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926672|gb|ACN69898.1|          ITRKKNQWDVLSNKFKSVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926674|gb|ACN69899.1|          ITRKKNQWDVLSNKFKSVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926676|gb|ACN69900.1|          ITRKKNQWDVLSNKFKSVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926678|gb|ACN69901.1|          ITRKKNQWDVLSNKFKSVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926680|gb|ACN69902.1|          ITRKKNQWDVLSNKFISVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926682|gb|ACN69903.1|          ITRKKNQWDVLSNKFKSVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926684|gb|ACN69904.1|          ITRKKNQWDVLSNKFISVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926686|gb|ACN69905.1|          ITRKKNQWDVLSNKFISVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926688|gb|ACN69906.1|          ITRKKNQWDVLSNKFISVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
```

```
gi|224926690|gb|ACN69907.1|         ITRKKNQWDVLSNKFKSVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926692|gb|ACN69908.1|         ITRKKNQWDVLSNKFISVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926694|gb|ACN69909.1|         VTRKKNQWDVLSNKFKSVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926696|gb|ACN69910.1|         ITRKKNQWDVLSNKFKSVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926698|gb|ACN69911.1|         ITRKKNQWDVLSNKFKSVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926700|gb|ACN69912.1|         VTRKKNQWDVLSNKFKSVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926702|gb|ACN69913.1|         ITRKKNQWDVLSNKFKSVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926704|gb|ACN69914.1|         ITRKKNQWDVLSNKFKSVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926706|gb|ACN69915.1|         ITRKKNQWDVLSNKFISVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926708|gb|ACN69916.1|         ITRKKNQWDVLSNKFKSVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926710|gb|ACN69917.1|         ITRKKNQWDVLSNKFISVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926714|gb|ACN69918.1|         ITRKKNQWDVLSNKFKSVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926716|gb|ACN69919.1|         ITRKKNQWDVLSNKFKSVKNAEKVQTAGIVTPYDILKQELDEFNEVALEN
gi|224926718|gb|ACN69920.1|         ITRKKNQWDVLSNKFISVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926720|gb|ACN69921.1|         ITRKKNQWDVLSNKFISVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926722|gb|ACN69922.1|         ITRKKNQWDVLSNKFKSVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926724|gb|ACN69923.1|         ITRKKNQWDVLSNKFISVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926726|gb|ACN69924.1|         ITRKKNQWDVLSNKFKSVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926728|gb|ACN69925.1|         ITRKKNQWDVLSNKFISVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926730|gb|ACN69926.1|         ITRKKNQWDVLSNKFISVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926732|gb|ACN69927.1|         ITRKKNQWDVLSNKFKSVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926734|gb|ACN69928.1|         ITRKKNQWDVLSNKFKSVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926736|gb|ACN69929.1|         ITRKKNQWDVLSNKFISVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926738|gb|ACN69930.1|         ITRKKNQWDVLSNKFKSVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926740|gb|ACN69931.1|         ITRKKNQWDVLSNKFKSVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926742|gb|ACN69932.1|         ITRKKNQWDVLSNKFKSVKNAEKVQTAGIVTPYDILKQELDEFIEVAFEN
gi|224926744|gb|ACN69933.1|         ITRKKNQWDVLSNKFISVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926746|gb|ACN69934.1|         ITRKKNQWDVLSNKFISVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926748|gb|ACN69935.1|         ITRKKNQWDVLSNKFKSVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926750|gb|ACN69936.1|         ITRKKNQWDVLSNKFKSVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926752|gb|ACN69937.1|         ITRKKNQWDVLSNKFISVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926754|gb|ACN69938.1|         ITRKKNQWDVLSNKFKSVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926756|gb|ACN69939.1|         ITRKKNQWDVLSNKFISVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926758|gb|ACN69940.1|         ITRKKNQWDVLSNKFKSVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926760|gb|ACN69941.1|         ITRKKNQWDVLSNKFKSVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926762|gb|ACN69942.1|         ITRKKNQWDVLSNKFKSVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926764|gb|ACN69943.1|         ITRKKNQWDVLSNKFKSVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926766|gb|ACN69944.1|         ITRKKNQWDVLSNKFKSVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926768|gb|ACN69945.1|         ITRKKNQWDVLSNKFISVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926770|gb|ACN69946.1|         ITRKKNQWDVLSNKFKSVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926772|gb|ACN69947.1|         ITRKKNQWDVLSNKFISVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926774|gb|ACN69948.1|         ITRKKNQWDVLSNKFISVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926776|gb|ACN69949.1|         ITRKKNQWDVLSNKFISVKNAEKVQTAGIVTPYDILKQELDEFNEMAFEN
gi|224926778|gb|ACN69950.1|         ITRKKNQWDVLSNKFISVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926780|gb|ACN69951.1|         ITRKKNQWDVLSNKFKSVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926782|gb|ACN69952.1|         ITRKKNQWDVLSNKFISVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926784|gb|ACN69953.1|         ITRKKNQWDVLSNKFKSVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926786|gb|ACN69954.1|         ITRKKNQWDVLSNKFISVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926788|gb|ACN69955.1|         ITRKKNQWDVLSNKFKSVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926790|gb|ACN69956.1|         ITRKKNQWDVLSNKFISVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926792|gb|ACN69957.1|         ITRKKNQWDVLSNKFISVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926794|gb|ACN69958.1|         ITRKKNQWDVLSNKFISVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926796|gb|ACN69959.1|         ITRKKNQWDVLSNKFKSVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926798|gb|ACN69960.1|         ITRKKNQWDVLSNKFKSVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926800|gb|ACN69961.1|         ITRKKNQWDVLSNKFISVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926802|gb|ACN69962.1|         ITRKKNQWDVLSNKFISVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926804|gb|ACN69963.1|         ITRKKNQWDVLSNKFKSVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926806|gb|ACN69964.1|         ITRKKNQWDVLSNKFKSVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926808|gb|ACN69965.1|         ITRKKNQWDVLSNKFISVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
```

792

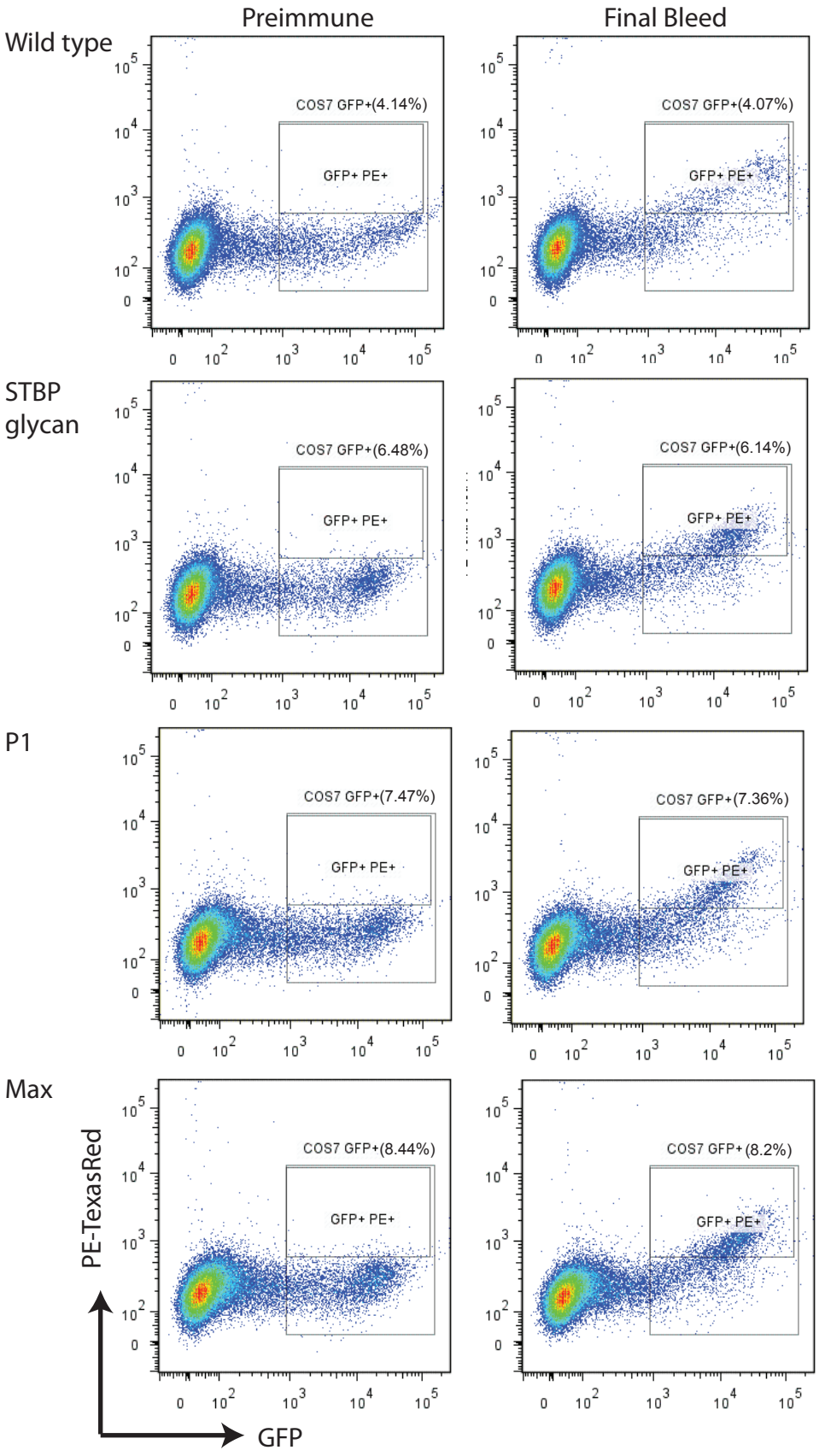
```
gi|224926810|gb|ACN69966.1|         ITRKKNQWDVLSNKFKSVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926812|gb|ACN69967.1|         ITRKKNQWDVLSNKFISVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926814|gb|ACN69968.1|         ITRKKNQWDVLSNKFISVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926816|gb|ACN69969.1|         ITRKKNQWDVLSNKFISVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926818|gb|ACN69970.1|         ITRKKNQWDVLSNKFISVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|224926820|gb|ACN69971.1|         ITRKKNQWDVLSNKFKSVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|213868281|gb|ACJ54187.1|         ITRKKNQWDVLSNKFISVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|213868283|gb|ACJ54188.1|         ITRKKNQWDVLSNKFISVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|213868285|gb|ACJ54189.1|         ITRKKNQWDVLSNKFISVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|213868287|gb|ACJ54190.1|         ITRKKNQWDVLSNKFKSVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|213868289|gb|ACJ54191.1|         ITRKKNQWDVLSNKFKSVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|213868291|gb|ACJ54192.1|         ITRKKNQWDVLSNKFKSVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|213868293|gb|ACJ54193.1|         ITRKKNQWDVLSNKFKSVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|213868295|gb|ACJ54194.1|         ITRKKNQWDVLSNKFKSVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|213868297|gb|ACJ54195.1|         ITRKKNQWDVLSNKFISVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|213868299|gb|ACJ54196.1|         ITRKKNQWDVLSNKFKSVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
gi|213868301|gb|ACJ54197.1|         ITRKKNQWDVLSNKFKSVKNAEKVQTAGIVTPYDILKQELDEFNEVAFEN
                                    :************** ****************

gi|11120233|gb|AAG30847.1|          EINKRDGAYIELCVCSVEEAKKNTQEVVT
gi|12407747|gb|AAG53634.1|          EINKRDGAYIELCVCSVEEAKKNTQEVVT
gi|18766659|gb|AAL79050.1|AF46      ----------------------------
gi|18766819|gb|AAL79130.1|AF46      ----------------------------
gi|63146462|gb|AAY34059.1|          ----------------------------
gi|63146616|gb|AAY34136.1|          ----------------------------
gi|11138983|gb|AAG31571.1|          EINKRDGAYIELCVCSVEEAKKNTQEVVT
gi|73697858|gb|AAZ81532.1|          EINKRDGAYIELCVCSVEEAKKNTQEVVT
gi|73697846|gb|AAZ81526.1|          EINKRDGAYIELCVCSVEEAKKNTQEVVT
gi|12004297|gb|AAG43989.1|          EINKRDGAYIELCVCSVEEAKKNTQEVVT
gi|12004299|gb|AAG43990.1|          EINKRDGAYIELCVCSVEEAKKNTQEVVT
gi|73697866|gb|AAZ81536.1|          EINKRDGAYIELCVCSVEEAKKNTQEVVT
gi|73697848|gb|AAZ81527.1|          EINKRDGAYIELCVCSVEEAKKNTQEVVT
gi|73697856|gb|AAZ81531.1|          EINKRDGAYIELCVCSVEEAKKNTQEVVT
gi|73697862|gb|AAZ81534.1|          EINKRDGAYIELCVCSVEEAKKNTQEVVT
gi|160276|gb|AAA63423.1|            EINKRDGAYIELCVCSVEEAKKNTQEVVT
gi|73697844|gb|AAZ81525.1|          EINKRDGAYIELCVCSVEEAKKNTQEVVT
gi|339961368|pdb|3RRC|A             EINKRDGAYIELCVCSVEEAKKNTQEVVT
gi|339961369|pdb|3RRC|B             EINKRDGAYIELCVCSVEEAKKNTQEVVT
gi|255914751|dbj|BAH96577.1|        EINKRDGAYIELCVCSVEEAKKNTQEVVT
gi|224926624|gb|ACN69874.1|         EINKRDGAYIELCV---------------
gi|224926626|gb|ACN69875.1|         EINKRDGAYIELCV---------------
gi|224926628|gb|ACN69876.1|         EINKRDGAYIELCV---------------
gi|224926630|gb|ACN69877.1|         EINKRDGAYIELCV---------------
gi|224926632|gb|ACN69878.1|         EINKRDGAYIELCV---------------
gi|224926634|gb|ACN69879.1|         EINKRDGAYIELCV---------------
gi|224926636|gb|ACN69880.1|         EINKRDGAYIELCV---------------
gi|224926638|gb|ACN69881.1|         EINKRDGAYIELCV---------------
gi|224926640|gb|ACN69882.1|         EINKRDGAYIELCV---------------
gi|224926642|gb|ACN69883.1|         EINKRDGAYIELCV---------------
gi|224926644|gb|ACN69884.1|         EINKRDGAYIELCV---------------
gi|224926646|gb|ACN69885.1|         EINKRDGAYIELCV---------------
gi|224926648|gb|ACN69886.1|         EINKRDGAYIELCV---------------
gi|224926650|gb|ACN69887.1|         EINKRDGAYIELCV---------------
gi|224926652|gb|ACN69888.1|         EINKRDGAYIELCV---------------
gi|224926654|gb|ACN69889.1|         EINKRDGAYIELCV---------------
gi|224926656|gb|ACN69890.1|         EINKRDGAYIELCV---------------
gi|224926658|gb|ACN69891.1|         EINKRDGAYIELCV---------------
gi|224926660|gb|ACN69892.1|         EINKRDGAYIELCV---------------
gi|224926662|gb|ACN69893.1|         EINKRDGAYIELCV---------------
```

```
gi|224926664|gb|ACN69894.1|         EINKRDGAYIELCV---------------
gi|224926666|gb|ACN69895.1|         EINKRDGAYIELCV---------------
gi|224926668|gb|ACN69896.1|         EINKRDGAYIELCV---------------
gi|224926670|gb|ACN69897.1|         EINKRDGAYIELCV---------------
gi|224926672|gb|ACN69898.1|         EINKRDGAYIELCV---------------
gi|224926674|gb|ACN69899.1|         EINKRDGAYIELCV---------------
gi|224926676|gb|ACN69900.1|         EINKRDGAYIELCV---------------
gi|224926678|gb|ACN69901.1|         EINKRDGAYIELCV---------------
gi|224926680|gb|ACN69902.1|         EINKRDGAYIELCV---------------
gi|224926682|gb|ACN69903.1|         EINKRDGAYIELCV---------------
gi|224926684|gb|ACN69904.1|         EINKRDGAYIELCV---------------
gi|224926686|gb|ACN69905.1|         EINKRDGAYIELCV---------------
gi|224926688|gb|ACN69906.1|         EINKRDGAYIELCV---------------
gi|224926690|gb|ACN69907.1|         EINKRDGAYIELCV---------------
gi|224926692|gb|ACN69908.1|         EINKRDGAYIELCV---------------
gi|224926694|gb|ACN69909.1|         EINKRDGAYIELCV---------------
gi|224926696|gb|ACN69910.1|         EINKRDGAYIELCV---------------
gi|224926698|gb|ACN69911.1|         EINKRDGAYIELCV---------------
gi|224926700|gb|ACN69912.1|         EINKRDGAYIELCV---------------
gi|224926702|gb|ACN69913.1|         EINKRDGAYIELCV---------------
gi|224926704|gb|ACN69914.1|         EINKRDGAYIELCV---------------
gi|224926706|gb|ACN69915.1|         EINKRDGAYIELCV---------------
gi|224926708|gb|ACN69916.1|         EINKRDGAYIELCV---------------
gi|224926710|gb|ACN69917.1|         EINKRDGAYIELCV---------------
gi|224926714|gb|ACN69918.1|         EINKRDGAYIELCV---------------
gi|224926716|gb|ACN69919.1|         EINKRDGAYIELCV---------------
gi|224926718|gb|ACN69920.1|         EINKRDGAYIELCV---------------
gi|224926720|gb|ACN69921.1|         EINKRDGAYIELCV---------------
gi|224926722|gb|ACN69922.1|         EINKRDGAYIELCV---------------
gi|224926724|gb|ACN69923.1|         EINKRDGAYIELCV---------------
gi|224926726|gb|ACN69924.1|         EINKRDGAYIELCV---------------
gi|224926728|gb|ACN69925.1|         EINKRDGAYIELCV---------------
gi|224926730|gb|ACN69926.1|         EINKRDGAYIELCV---------------
gi|224926732|gb|ACN69927.1|         EINKRDGAYIELCV---------------
gi|224926734|gb|ACN69928.1|         EINKRDGAYIELCV---------------
gi|224926736|gb|ACN69929.1|         EINKRDGAYIELCV---------------
gi|224926738|gb|ACN69930.1|         EINKRDGAYIELCV---------------
gi|224926740|gb|ACN69931.1|         EINKRDGAYIELCV---------------
gi|224926742|gb|ACN69932.1|         EINKRDGAYIELCV---------------
gi|224926744|gb|ACN69933.1|         EINKRDGAYIELCV---------------
gi|224926746|gb|ACN69934.1|         EINKRDGAYIELCV---------------
gi|224926748|gb|ACN69935.1|         EINKRDGAYIELCV---------------
gi|224926750|gb|ACN69936.1|         EINKRDGAYIELCV---------------
gi|224926752|gb|ACN69937.1|         EINKRDGAYIELCV---------------
gi|224926754|gb|ACN69938.1|         EINKRDGAYIELCV---------------
gi|224926756|gb|ACN69939.1|         EINKRDGAYIELCV---------------
gi|224926758|gb|ACN69940.1|         EINKRDGAYIELCV---------------
gi|224926760|gb|ACN69941.1|         EINKRDGAYIELCV---------------
gi|224926762|gb|ACN69942.1|         EINKRDGAYIELCV---------------
gi|224926764|gb|ACN69943.1|         EINKRDGAYIELCV---------------
gi|224926766|gb|ACN69944.1|         EINKRDGAYIELCV---------------
gi|224926768|gb|ACN69945.1|         EINKRDGAYIELCV---------------
gi|224926770|gb|ACN69946.1|         EINKRDGAYIELCV---------------
gi|224926772|gb|ACN69947.1|         EINKRDGAYIELCV---------------
gi|224926774|gb|ACN69948.1|         EINKRDGAYIELCV---------------
gi|224926776|gb|ACN69949.1|         EINKRDGAYIELCV---------------
gi|224926778|gb|ACN69950.1|         EINKRDGAYIELCV---------------
gi|224926780|gb|ACN69951.1|         EINKRDGAYIELCV---------------
gi|224926782|gb|ACN69952.1|         EINKRDGAYIELCV---------------
```

794

```
gi|224926784|gb|ACN69953.1|         EINKRDGAYIELCV---------------
gi|224926786|gb|ACN69954.1|         EINKRDGAYIELCV---------------
gi|224926788|gb|ACN69955.1|         EINKRDGAYIELCV---------------
gi|224926790|gb|ACN69956.1|         EINKRDGAYIELCV---------------
gi|224926792|gb|ACN69957.1|         EINKRDGAYIELCV---------------
gi|224926794|gb|ACN69958.1|         EINKRDGAYIELCV---------------
gi|224926796|gb|ACN69959.1|         EINKRDGAYIELCV---------------
gi|224926798|gb|ACN69960.1|         EINKRDGAYSELCV---------------
gi|224926800|gb|ACN69961.1|         EINKRDGAYIELCV---------------
gi|224926802|gb|ACN69962.1|         EINKRDGAYIELCV---------------
gi|224926804|gb|ACN69963.1|         EINKRDGAYIELCV---------------
gi|224926806|gb|ACN69964.1|         EINKRDGAYIELCV---------------
gi|224926808|gb|ACN69965.1|         EINKRDGAYIELCV---------------
gi|224926810|gb|ACN69966.1|         EINKRDGAYIELCV---------------
gi|224926812|gb|ACN69967.1|         EINKRDGAYIELCV---------------
gi|224926814|gb|ACN69968.1|         EINKRDGAYIELCV---------------
gi|224926816|gb|ACN69969.1|         EINKRDGAYIELCV---------------
gi|224926818|gb|ACN69970.1|         EINKRDGAYIELCV---------------
gi|224926820|gb|ACN69971.1|         EINKRDGAYIELCV---------------
gi|213868281|gb|ACJ54187.1|         EINKRDGAYIELCVCSVEEAKKNTQEVVT
gi|213868283|gb|ACJ54188.1|         EINKRDGAYIELCVCSVEEAKKNTQEVVT
gi|213868285|gb|ACJ54189.1|         EINKRDGAYIELCVCSVEEAKKNTQEVVT
gi|213868287|gb|ACJ54190.1|         EINKRDGAYIELCVCSVEEAKKNTQEVVT
gi|213868289|gb|ACJ54191.1|         EINKRDGAYIELCVCSVEEAKKNTQEVVT
gi|213868291|gb|ACJ54192.1|         EINKRDGAYIELCVCSVEEAKKNTQEVVT
gi|213868293|gb|ACJ54193.1|         EINKRDGAYIELCVCSVEEAKKNTQEVVT
gi|213868295|gb|ACJ54194.1|         EINKRDGAYIELCVCSVEEAKKNTQEVVT
gi|213868297|gb|ACJ54195.1|         EINKRDGAYIELCVCSVEEAKKNTQEVVT
gi|213868299|gb|ACJ54196.1|         EINKRDGAYIELCVCSVEEAKKNTQEVVT
gi|213868301|gb|ACJ54197.1|         EINKRDGAYIELCVCSVEEAKKNTQEVVT
```

Figure S3. Transfection of COS-7 cells with PvDBPII-GFP fusion proteins. The gating strategy for GFP and anti-PvDBPII doubly positive transfected COS-7 cells is shown. The gate for GFP-positive cells was set by comparison to untransfected cells. The gate for anti-PvDBPII-positive (PE-Texas Red) was set by comparison to cells labeled with secondary antibody alone.


Preimmune
Final Bleed
Wild type
STBP glycan
P1
Max
COS7 GFP+(4.14%)
COS7 GFP+(4.07%)
COS7 GFP+(6.48%)
COS7 GFP+(6.14%)
COS7 GFP+(7.47%)
COS7 GFP+(7.36%)
COS7 GFP+(8.44%)
COS7 GFP+(8.2%)
GFP+ PE+
PE-TexasRed
GFP

800

Figure S4. Western blot of DBPII glycosylation variants expressed in COS-7 cells with or without PNGaseF treatment. COS-7 cells transfected with recombinant DBPII glycosylation variants were lysed 48 h post transfection and immunoprecipitated with anti-GFP agarose resin. Half of the sample was subjected to PNGaseF treatment and the other half was untreated to observe glycosylation modifications.

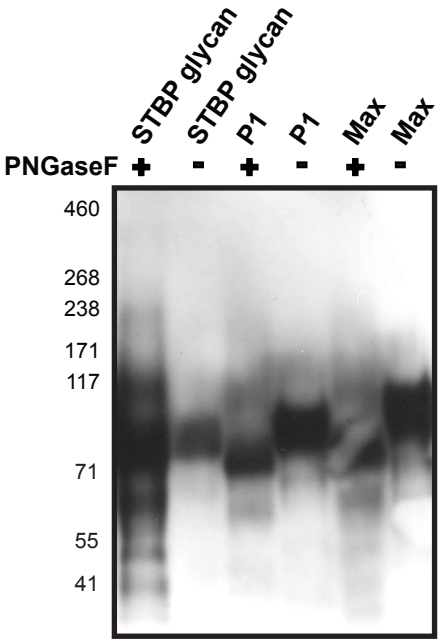


Figure S5. Inhibition of PvDBPII binding to DARC in different assay formats. Mice were immunized with wild-type DBPII protein produced in *E. coli* or HEK293 cells. Antibody inhibition of PvDBPII-DARC interaction in COS-7-RBC binding inhibition assay (A) and yeast display binding inhibition assay format (B).

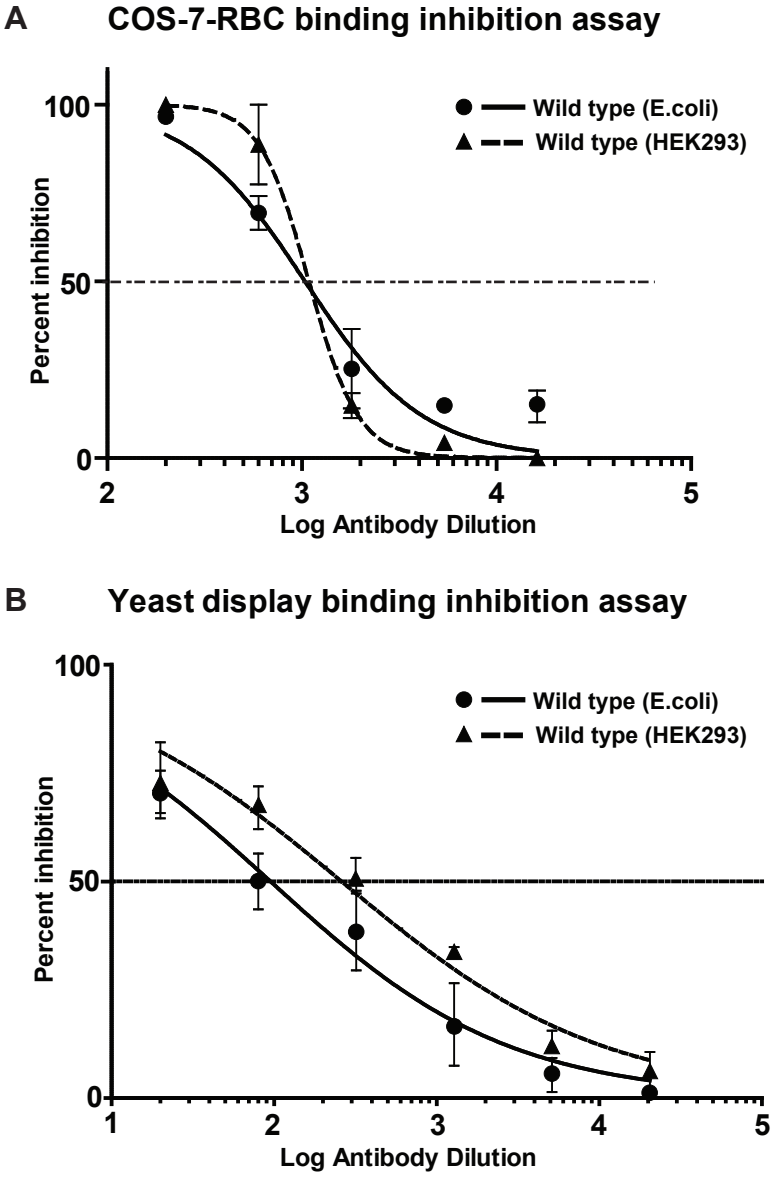


Figure S6. Yeast display antibody binding inhibition assay. Histograms showing DARC-Fc binding to yeast. (A) Uninduced yeast with no PvDBPII surface expression plus wild-type DARC-Fc (negative control). (B) Induced yeast with wild-type DARC-Fc (positive control). (C) Induced yeast with an inactive DARC-Fc mutant (negative control).

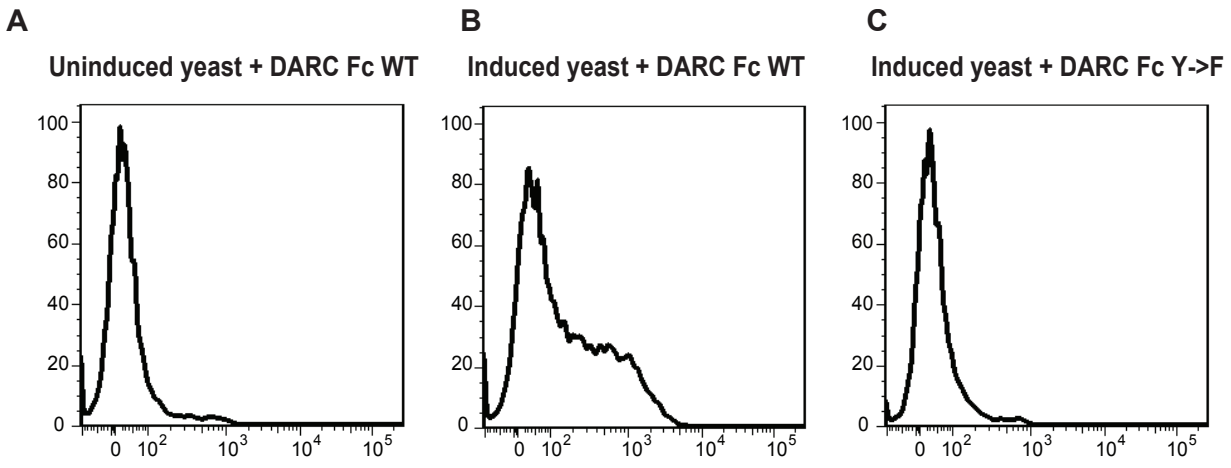


TABLE S1. Immunization schedule

Immunization	Immunogen	Group	Animal	No. of animals	No. of Immunizations		Dose (μg)	
					DNA	Protein	DNA	Protein
Protein[1]	Wt (*E. coli* refolded)	1	Mouse	10		4		10
Protein[1]	Wt (HEK293)	2	Mouse	10		4		5
DNA[2],Protein[1]	Wt (HEK293)	3	Mouse	10	4	2	20	2.5
DNA[2],Protein[1]	STBP glycan	4	Mouse	10	4	2	20	2.5
DNA[2],Protein[1]	P1	5	Mouse	10	4	2	20	2.5
DNA[2],Protein[1]	MAX	6	Mouse	10	4	2	20	2.5

[1,] immunization was intramuscular

[2,] immunization was by DNA electroporation

815

Table S2. Adhesion of PvDBP mutants to DARC

Mutation	Location[1]	Orientation of amino acid[2]	Reduction on DARC binding[6]	Reference
Δ 198-216	SD1		none	Bolton et al. Virology J 2011
Δ 198-247	SD1		major	Bolton et al. Virology J 2011
Δ198-216, Δ248-522	SD1, SD2, & SD3		major	Bolton et al. Virology J 2011
Δ 216-247	SD1		major	Bolton et al. Virology J 2011
K220A,R223A,R226A	SD1	surface	major	Bolton et al. Virology J 2011
K222A	SD1	surface (^)[3]	none	Bolton et al. Virology J 2011
T232N	SD1	surface (*)[4]	none	this study
K234S	SD1	surface (^)	none	this study
Δ 248-522	SD2 & SD3		major	Bolton et al. Virology J 2011
E249R	SD1	surface (^)	none	Batchelor et al. NSMB 2011
E249R,R274E	SD1 &SD2	surface (^)	moderate	Batchelor et al. NSMB 2011
L250A	SD1	buried	none	VanBuskirk et al. PNAS 2004
T251A	SD1	surface (*)	none	VanBuskirk et al. PNAS 2004
V254A	SD1	No info[5]	minor	VanBuskirk et al. PNAS 2004
N255A	SD1	No info	minor	VanBuskirk et al. PNAS 2004
N256A	SD1	No info	minor	VanBuskirk et al. PNAS 2004
T257A	SD1	No info	minor	VanBuskirk et al. PNAS 2004
D258A	SD1	No info	moderate	VanBuskirk et al. PNAS 2004
T259A	SD1	No info	moderate	VanBuskirk et al. PNAS 2004
N260A	SD1	No info	major	VanBuskirk et al. PNAS 2004
F261A	SD1	No info	major	VanBuskirk et al. PNAS 2004
D264N	SD1	No info	none	this study
T266A	SD2	surface (*)	none	VanBuskirk et al. PNAS 2004
F267A	SD2	surface (*)	none	VanBuskirk et al. PNAS 2004
L270A	SD2	surface (*)	none	VanBuskirk et al. PNAS 2004
Y271A	SD2	surface (^)	none	VanBuskirk et al. PNAS 2004
L272A	SD2	buried	major	VanBuskirk et al. PNAS 2004
K273A	SD2	surface (^)	major	VanBuskirk et al. PNAS 2004
K273A	SD2	surface (^)	moderate	Batchelor et al. NSMB 2011
R274A	SD2	surface (^)	none	Hans et al. Mol Micro 2005
R274A	SD2	surface (^)	moderate	Batchelor et al. NSMB 2011
R274E	SD2	surface (^)	major	Batchelor et al. NSMB 2011
K275A	SD2	surface (^)	moderate	VanBuskirk et al. PNAS 2004
L276A	SD2	buried	moderate	VanBuskirk et al. PNAS 2004
I277A	SD2	surface (*)	moderate	VanBuskirk et al. PNAS 2004
L290A	SD2	buried	none	VanBuskirk et al. PNAS 2004
N291A	SD2	surface (*)	moderate	Hans et al. Mol Micro 2005
N291A	SD2	surface (*)	moderate	VanBuskirk et al. PNAS 2004
N292A	SD2	surface (^)	none	VanBuskirk et al. PNAS 2004
Y293A	SD2	surface (^)	none	VanBuskirk et al. PNAS 2004
Y293A	SD2	surface (^)	minor	Hans et al. Mol Micro 2005
R294A	SD2	surface (^)	none	VanBuskirk et al. PNAS 2004

816

Y295A	SD2	surface (^)	major	VanBuskirk et al. PNAS 2004
N296A	SD2	surface (*)	moderate	VanBuskirk et al. PNAS 2004
K297A	SD2	surface (^)	major	VanBuskirk et al. PNAS 2004
F299A	SD2	buried	major	VanBuskirk et al. PNAS 2004
F299A	SD2	buried	major	Hans et al. Mol Micro 2005
D302A	SD2	buried	minor	VanBuskirk et al. PNAS 2004
I303A	SD2	buried	major	VanBuskirk et al. PNAS 2004
R304A	SD2	surface (^)	major	VanBuskirk et al. PNAS 2004
V327A	SD2	surface (^)	none	VanBuskirk et al. PNAS 2004
V328A	SD2	buried	none	VanBuskirk et al. PNAS 2004
E329A	SD2	surface (*)	major	VanBuskirk et al. PNAS 2004
N330A	SD2	surface (^)	major	VanBuskirk et al. PNAS 2004
D339A	SD2	surface (^)	minor	Hans et al. Mol Micro 2005
E340A	SD2	surface (^)	none	Hans et al. Mol Micro 2005
E340A	SD2	surface (^)	none	VanBuskirk et al. PNAS 2004
K341A	SD2	surface (^)	none	VanBuskirk et al. PNAS 2004
K341N	SD2	surface (^)	none	this study
Q343A	SD2	surface (^)	none	VanBuskirk et al. PNAS 2004
Q343S	SD2	surface (^)	none	this study
Q344A	SD2	surface (^)	none	VanBuskirk et al. PNAS 2004
Q344A	SD2	surface (^)	none	Hans et al. Mol Micro 2005
R345A	SD2	surface (^)	none	VanBuskirk et al. PNAS 2004
K347A	SD2	surface (^)	none	VanBuskirk et al. PNAS 2004
Q348A	SD2	surface (^)	none	VanBuskirk et al. PNAS 2004
Q348A	SD2	surface (^)	none	Hans et al. Mol Micro 2005
W350A	SD2	buried	major	VanBuskirk et al. PNAS 2004
N351A	SD2	surface (^)	minor	VanBuskirk et al. PNAS 2004
N351A	SD2	surface (^)	none	Batchelor et al. NSMB 2011
E352A	SD2	surface (^)	minor	VanBuskirk et al. PNAS 2004
K354A	SD2	surface (^)	none	Batchelor et al. NSMB 2011
Q356A	SD2	surface (^)	none	Batchelor et al. NSMB 2011
Y363A	SD2	surface (^)	major	Hans et al. Mol Micro 2005
S364A	SD2	surface (^)	moderate	VanBuskirk et al. PNAS 2004
V365A	SD2	surface (*)	moderate	VanBuskirk et al. PNAS 2004
K366A	SD2	No info	major	VanBuskirk et al. PNAS 2004
K367A,K370A	SD2	No info	none	Bolton et al. Virology J 2011
K367A	SD2	No info	major	VanBuskirk et al. PNAS 2004
R368A	SD2	No info	moderate	VanBuskirk et al. PNAS 2004
L369A	SD2	No info	major	VanBuskirk et al. PNAS 2004
K370A	SD2	No info	none	VanBuskirk et al. PNAS 2004
F373A	SD2	No info	major	Hans et al. Mol Micro 2005
I374N,I376T	SD2	No info	major	this study
I376A	SD2	No info	major	Hans et al. Mol Micro 2005
C377A	SD2	surface (*)	none	VanBuskirk et al. PNAS 2004
N384A	SD2	surface (^)	none	Batchelor et al. NSMB 2011
Q388A	SD3	buried	moderate	VanBuskirk et al. PNAS 2004
Y390A	SD3	buried	moderate	VanBuskirk et al. PNAS 2004

817

R391A	SD3	buried	minor	VanBuskirk et al. PNAS 2004
R394A	SD3	surface (^)	moderate	VanBuskirk et al. PNAS 2004
E395A	SD3	buried	moderate	VanBuskirk et al. PNAS 2004
W396A	SD3	buried	moderate	VanBuskirk et al. PNAS 2004
G397A	SD3	buried	minor	VanBuskirk et al. PNAS 2004
R398A	SD3	surface (^)	none	Hans et al. Mol Micro 2005
D399A	SD3	buried	minor	VanBuskirk et al. PNAS 2004
Y400A	SD3	buried	minor	VanBuskirk et al. PNAS 2004
V401A	SD3	surface (^)	minor	VanBuskirk et al. PNAS 2004
S402A	SD3	surface (^)	none	VanBuskirk et al. PNAS 2004
E403A	SD3	surface (^)	none	VanBuskirk et al. PNAS 2004
L404A	SD3	buried	none	VanBuskirk et al. PNAS 2004
P405A	SD3	surface (^)	none	VanBuskirk et al. PNAS 2004
T406A	SD3	surface (^)	none	VanBuskirk et al. PNAS 2004
E407A	SD3	surface (*)	none	VanBuskirk et al. PNAS 2004
V408A	SD3	buried	none	VanBuskirk et al. PNAS 2004
K412N	SD3	surface (^)	none	this study
E413A	SD3	surface (^)	none	VanBuskirk et al. PNAS 2004
K414A	SD3	surface (^)	none	VanBuskirk et al. PNAS 2004
K414S	SD3	surface (^)	none	this study
I419A	SD3	surface (^)	none	VanBuskirk et al. PNAS 2004
N420A	SD3	surface (*)	none	VanBuskirk et al. PNAS 2004
K425A	SD3	surface (^)	none	VanBuskirk et al. PNAS 2004
V426A	SD3	surface (*)	none	VanBuskirk et al. PNAS 2004
E464S	SD3	surface (^)	none	this study
Q467N	SD3	surface (^)	none	this study
A469T	SD3	surface (^)	none	this study
V488T	SD3	surface (^)	none	this study
R497S	SD3	surface (^)	none	this study
D264N,E464S,V488T	SD2 & SD3	surface	none	this study
T232N,K234S,D264N,K341N, Q343S,K412N,K414S,Q467N, A469T,V488T,R497S	SD1, SD2, & SD3	surface	none	this study

[1,] subdomain designations according to Singh et al. Nature 2006

[2,]designation of orientation of amino acid is according to 3RRC.pdb

[3,] ^ means surface exposed "R" group

[4,] * means buried amino acid or the "R" group is not directed outward

[5,] no info means no structural information is available due to the absence of density in PvDBPII crystal structure 3RRC.pdb

[6,] reduction of DARC binding is as describe in Hans et al., Bolton et al., Batchelor et al., and this study where none is 70-100% binding, minor is 50-70% binding, moderate is 30-50% binding, and major is less than 30% binding. Reduction of DARC binding for VanBuskirk et al. is as described in their manuscript where none is 90-100% binding, minor effect is 50-90% binding, moderate is 5-50% binding, and major is 0-5% binding

818

Table S3. IC50 of DBPII glycosylation variants.

Immunogen	COS-7-RBC binding inhibition assay		Yeast display binding inhibition assay		Effect of DARC phenotype			
					FyA		FyB	
	IC50[1]	SD[2]	IC50	SD	IC50	SD	IC50	SD
Wild type (HEK293)[3]	2589	8.03	32.71	21.45	1203	34.85	442.7	33.6
STBP glycan[3]	2787	13.82	34.12	23.32	435.9	32.67	408	28.92
P1[3]	2837	10.11	199.4	31.01	1237	40.14	459.7	42.91
Max[3]	2110	8.53	668.6	27.5	926.7	35.72	273.4	35.31

[1,] plasma dilution at 50% inhibition

[2,] standard deviation

[3,] DNA immunization,protein boost

TABLE 1. PvDBPII recombinant proteins

Immunogen	No. of glycosylation	N-glycosylation sites	Molecular mass (kDa)	Percent Binding
Wild type (HEK293)	3	N255, N351, N420	49	97
STBP glycan	4	N255, N351, N374, N420	51	4
P1	6	N255, N264, N351, N420, N462, N486	55	75
Max	11	N232, N255, N264, N341, N351, N412, N420, N462, N467, N486, N495	65	88