Table of Contents

TABLE OF FIGURES

TABLE OF FIGURES

ABSTRACT

Certain elite controllers have very potent broadly neutralizing antibodies that neutralize a range of HIV viruses by targeting certain conserved regions on the virus. Designing immunogens to elicit such broadly neutralizing antibodies capable of neutralizing all HIV variants and types that are found globally has proven extremely challenging. This finding has not only provided novel strategies of rational vaccine design but has also led to the emergence of epitope driven vaccines.

In the same direction we performed the computational grafting of discontinuous Bcell epitopes of gp120 on suitable scaffolds. 431 scaffolds were screened for sites capable of carrying the epitopes of interest. Out of the 431 candidates, four structures were the successfully designed which would be used by the wet lab experimentalists for further work and analysis.

The main idea behind the work is that given the appropriate immunogen, it should be possible to elicit these broadly neutralizing antibody responses by vaccination.

AIM

Computational designing of HIV immunogens

OBJECTIVES
1. Epitope prediction using differnt servers and generation of a consensus.
2. Computational alanine scanning of Ag-Ab complexes.
3. Computational grafting of discontinuous epitope on suitable scaffolds.
4. Validation by comparing the interactions of the actual complex with the designed complexes.

INTRODUCTION

The first cases of AIDS in Africa were described in the early 1980s where it affected heterosexual populations in a band of equatorial countries including Zaire (now Democratic Republic of Congo), Rwanda, and Uganda (1). It was initially known as "slim disease" because it was associated with diarrhea and extensive weight loss. The spread of HIV-1 to southern African countries occurred later where detection in the general population of, for example, South Africa was only observed in the late 1980s (3). Whereas the epidemic in Africa has always been driven by heterosexual transmission, in other regions such as Latin America and the Caribbean, the HIV epidemic was also associated with injection drug use (IDU) and men who have sex with men (MSM). By the year 2000, the Caribbean had the second highest prevalence of HIV-1-infected adult population outside sub-Saharan Africa (4).

In Asia, HIV/AIDS was first reported in Thailand, India, and China around the mid to late 1980s (5,6). Since the early 1990s, HIV incidences in excess of 10% were documented among female sex workers (FSW) in Thailand and India, as well as among IDUs in countries near the "golden triangle" region (where the borders of Thailand, Myanmar, and Laos meet) (7;8;9;10;11). The epidemic then spread to more countries and populations in Asia through IDU, heterosexual transmission, and, at a later time, through MSM contacts, as well as mother-to-child transmission.

The major attack of the epidemic globally is carried by the low- to middle-income countries, which provide the virus with a productive and fertile environment to spread owing to a combination of factors including poor socioeconomic conditions, lack of access to health care, economic or political displacement of communities, and, to a certain extent, cultural practices and gender inequalities. As per the UNAIDS 2010 statistics 91% of the world's total HIV-1-infected population, numbering 30 million people, resided in low- to middle-income countries in 2009. Being the worst hit is the sub-Saharan Africa, which holds about three-quarters of the world's HIV-1-infected individuals from low and middle-income countries and more than two-thirds of the world's total. Other regions with high HIV/AIDS burdens include South and Southeast Asia, Eastern Europe and Central Asia, and Central and South America, which, respectively, account for approximately 12%, 4%, and

4% of the world's HIV-1-infected individuals. Low- to middle-income countries also hold the heaviest burden of drug use–related HIV infections. The three regions with both the highest numbers and the highest prevalence of HIV-positive IDUs are all low- to middle income regions: Latin America, with 580,000 HIV-infected IDUs (a prevalence of 29%), Eastern Europe with 940,000 HIV-infected IDUs (a prevalence of 27%), and East and Southeast Asia with 661,000 HIV-infected IDUs (a prevalence of 17%) (The HIV epidemic: low to middle income countries). These associations also implies that the fight against drug trafficking and abuse in these regions is inextricably linked with the fight against HIV.

Molecular epidemiology

Genetically and antigenically two different types of HIV have been distinguished. While HIV-1 is more common and contributes to the majority of HIV infections worldwide, HIV-2 which is transmitted in a similar way and is clinically similar to HIV-1, progresses slowly and is limited to the regions of West Africa. HIV-1 and HIV-2 both have arisen from Simian immunodeficiency virus SIV harboring two different natural hosts, chimpanzees and mangabeys respectively.

HIV-2 is closely related to the SIV found in West Africa.

There are three sub-groups of HIV-1, M (main or major), N (new) and O (outlier). The pandemic today is caused by HIV group M viruses and currently circulating viruses can be categorized into nine distinct clusters, referred to as subtypes or clades (subtypes A–D, F–H, J, and K), two of which have evolved further into distinct "sub-subtype" lineages known as A1 and A2 for subtype A viruses, and F1 and F2 for subtype F viruses. In addition, lineages have evolved comprising viruses that have been derived through genetic recombination between viruses of different subtypes, which arose as a result of simultaneous infection of individuals by two different subtypes. Whereas countless unique recombinant viruses have emerged in areas where two or more subtypes cocirculate, when recombinant viruses are discovered to have spread to form their own subepidemics, they are referred to as circulating recombinant forms. There are currently 48 circulating recombinant forms, or CRFs named CRF01 through CRF48 (12). Type O HIV-1 is mostly found in Cameroon and Gabon while the rare N sub-group is also found in Cameroon. It is

very likely that SIVcpz infected humans on separate occasions to give rise to the three sub-groups.
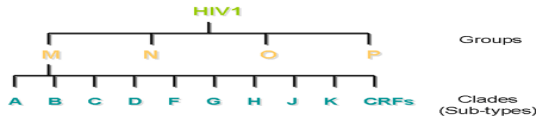
Globally, subtype C is now the most successful of the HIV-1M lineages and accounts for 50% of infections at present, whereas subtypes A and B each account for over 10% of worldwide HIV infections. Subtypes D, G, CRF01_AE, and CRF02_AG account for only between 2% and 6% each. Subtypes F, H, J, K, other CRFs, and all other unclassified recombinant forms individually make only a minor contribution to the global HIV population (1% each) but together account for the remaining 15% of worldwide HIV infections (13).

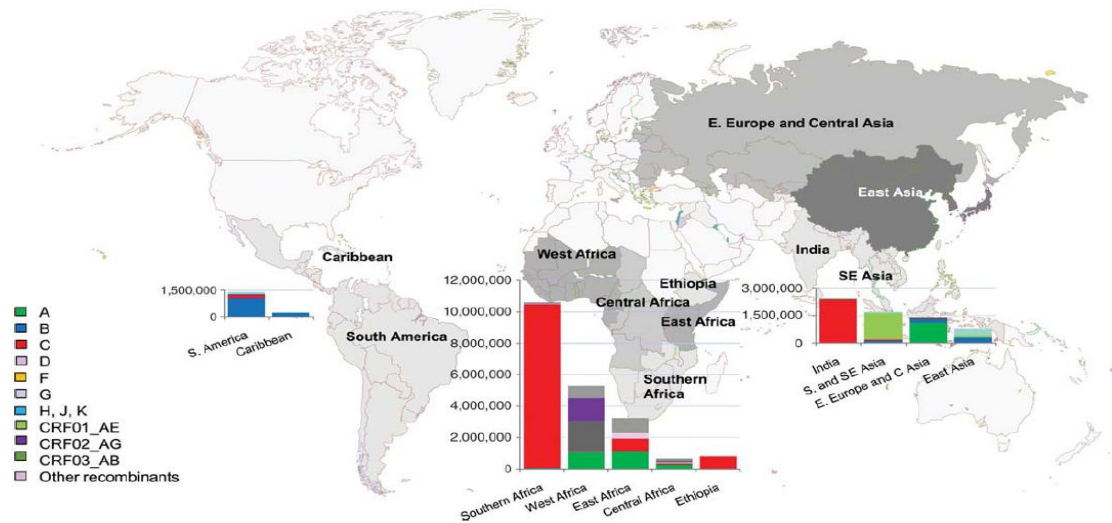**Figure 1: Regional distribution of HIV-1 subtypes and circulating recombinant forms in low- and middle-income countries. Regions comprising different countries are colored in different shades of gray. The size of the bar is proportional to the size of the epidemic (UNAIDS 2010) with the proportion of the subtypes contributing to each epidemic (as calculated by Hemelaar et al.2006) illustrated as a percentage of the bar.**
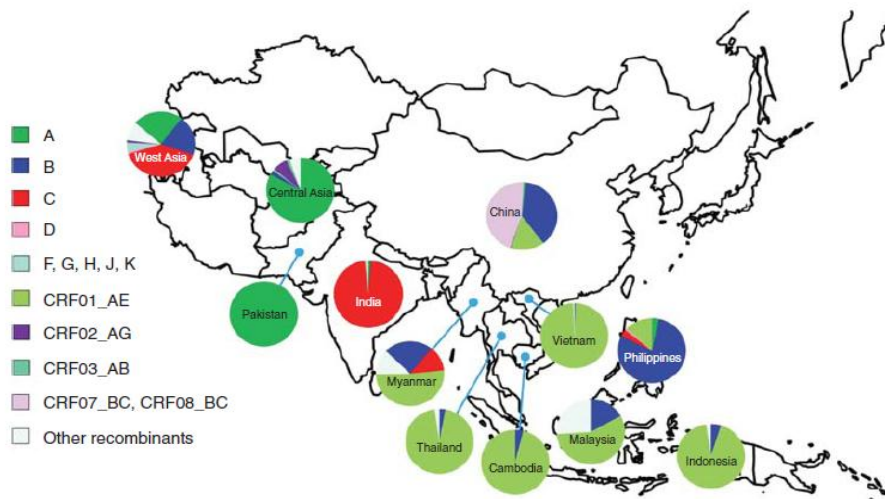
Figure 2: Regional distribution of HIV-1 subtypes and recombinants in Asia. The data of Malaysia, Indonesia, Philippines, Pakistan, Central Asia (including Afghanistan, Kazakhstan, and Uzbekistan) and West Asia (including Saudi Arabia, Iran and Israel) are from the Los Alamos HIV databases (http://www.hiv.lanl.gov/). The data of China, Japan, India, Myanmar, Vietnam and Thailand are from former studies in the region and estimates based on the unpublished data of research projects (Y Shao. pers.comm.).

Gp-120

The glycoprotein gp120 subunit is a major target for neutralizing antibodies and an important part of the envelope spikes that protrudes from the surface of HIV-1. However, immunization with recombinant gp120 does not elicit neutralizing antibodies against multiple HIV-1 isolates (broadly neutralizing antibodies), and gp120 failed to demonstrate vaccine efficacy in recent clinical trials. Immunization with recombinant monomeric gp120 elicited antibodies that could neutralize viruses that were neutralization-sensitive following extensive passage in immortalized T cell line cultures, but not viruses grown in limited passage in peripheral blood mononuclear cells (PBMCs) (14). The latter are more representative of circulating "primary" viruses. The general inability of recombinant gp120 to elicit cross-neutralizing (or broadly neutralizing) antibodies to primary viruses has sharpened interest in understanding the differences between the structure of gp120 in its recombinant form and its structure in the context of the envelope spike. Plethora of evidence suggests that the gp120 conformation in the context of the envelope spike is the crucial structure recognized by neutralizing antibodies (15-21).

The HIV envelope spike is a complex between gp120 and gp41 (22). The gp120 unit mediates the initial attachment of the virus to the target cell, whereas gp41 is required for the fusion of virus and target cell membranes. During HIV infection, the viral envelope spike is first synthesized as a single polypeptide precursor (22) and then subsequently oligomerizes and undergoes extensive glycosylation within the Golgi. The glycosylation process imparts conformational stability and ensures proper folding of the envelope glycoprotein (23); glycosylation primarily involves the attachment of N-linked high mannose-type oligosaccharides to the protein backbone. As the glycoprotein is transported through the Golgi, accessible glycan moieties are trimmed and modified by various cellular enzymes (22). These modifications generate so-called complex-type oligosaccharides; glycans that are relatively inaccessible to modifying enzymes remain as high-mannose type glycans (24). The resulting glycoprotein, having a molecular mass of ~160 kDa, is further cleaved in the trans-Golgi network by furin or equivalent endoproteases into gp120 and gp41 (22). The gp120-gp41 complexes, which remain associated through weak noncovalent interactions, are initially expressed at the surface of infected cells. During the HIV budding process, the gp120-gp41 complexes are then incorporated into the virus envelope and displayed on its surface as viral spikes (22).

The functional part of the envelope spike is a heterodimeric trimer complex of gp120 and gp41. The virion surface bears both the functional trimmers as well as other nonfunctional forms of the envelope. These nonfunctional envelope entities may be monomers, dimers, or tetramers and could possibly arise as the result of (*a*) the dissociation of functional gp120-gp41 complexes, which could perhaps cause gp120 to be shed from the viral surface, or (*b*) inefficient trimerization of the spike in the Golgi (22, 25, 26).

Based on comparative sequence analyses, gp120 has five conserved (C1–C5) and five variable (V1–V5) segments (27, 28). The C1 and C5 regions were reasoned to be the main areas on gp120 for contact with gp41, as these regions are accessible to antibody on monomeric gp120 but not on gp120-gp41 complexes (29, 30, 34). Other conserved segments C2, C3, and C4 regions were suggested to form a buried, relatively hydrophobic core within the gp120 molecule (30, 31). This gp120 core holds in it several discontinuous

neutralizing antibody epitopes that overlap the binding sites for CD4, the primary HIV receptor, and the coreceptor (30–33).

In contrast to the conserved regions, the variable regions (in particular, V1, V2, and V3) were argued to be well exposed on the surface of monomeric gp120 (30). Deletion of V1/V2 and V3 causes an increase in the binding affinity of antibodies to epitopes that overlap the binding sites for CD4 and the coreceptor, which suggests that these variable regions may shield conserved epitopes from efficient antibody recognition (35–38). For the V4 and V5 variable regions, no definitive role has been ascribed; deletion of the V4 region though has been shown to disrupt gp160 folding (31, 37), V4 also seems to tolerate insertion of foreign antibody epitopes (39).

The CD4 binding site: The binding site for CD4 on the liganded gp120 structure is formed by the interface between the inner domain, bridging sheet, and outer domain (42, 44). At the center of this interface lies a hydrophobic cavity that has been dubbed the Phe43 cavity (42). However, most of the CD4 contact residues are located on the outer domain of the liganded HIV-1 gp120 structures and form a contiguous binding region.

The co receptor binding site: The region that is important for the interaction with the β-chemokine receptor CCR5 has been mapped to residues in the bridging sheet and near the V3 stem (40, 41). These residues lie close together on the liganded HIV-1 gp120 structure, but the equivalent residues on the unliganded SIV gp120 structure are separated into two areas (**Figure *c,d***) (14). These differences are consistent with the notion that CD4 binding is required to lock these areas into a contiguous binding site. The fact that the coreceptor site is not presented until after CD4 binding suggests that the site may be susceptible to antibody recognition. Several studies have shown that HIV strains that do not require CD4 for entry are, in fact, highly sensitive to antibody neutralization (51–54). Owing to neutralizing antibody–driven selection pressure in vivo, the prevalence of such viruses during infection is likely low. However, in the absence of circulating neutralizing antibodies, e.g., in the central nervous system, such viruses may occur more frequently (54, 55).
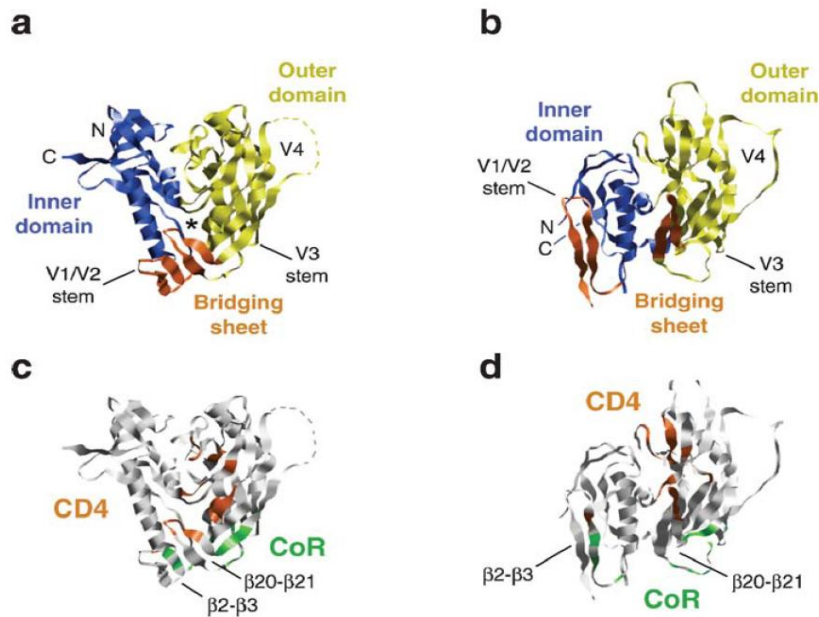
Figure 3: Comparison of the crystal structures of HIV-I and SIV gp120 core. (a) Structure of the CD4-liganded HIV-I gp 120 core (HXB2), viewed from the perspective of CD4. The gp120 inner domain (blue), outer domain (yellow), and bridging sheet (orange) are shown. The locations of various gp120 regions are also denoted. (b) Structure of the unliganded SIV gp120 core (Protein Data Bank ID 2BF1), viewed from the perspective CD4 as in (a). (c) HIV core gp120 in same ()tient:01°n as (a), depicting CD4 contact residues (orange) and residues that influence coreceptor binding (green). (d) SIV gp120 core in thc same, orientation as in (b). Colored according to the scheme for HIV gp120 in (c)

<u>Viral evasion of immune response</u>

1. Previous studies had suggested that the CoRbs is masked by the V1/V2 variable regions (35, 36); anti-CoRbs antibodies typically require CD4 for high-affinity binding to wild-type gp120, but removal of V1/V2 allows these antibodies to bind their epitopes in the absence of CD4 with the same high affinity as when binding to wild-type gp120 in the presence of CD4. The locations of the V1/V2 stem in the CD4-liganded and -unliganded gp120 structures relative to the coreceptor site suggests that V1/V2 may not be close enough to obstruct the CoRbs within the same gp120 molecule on the virus (42, 43). However, oligomeric modeling of the CD4- liganded structure does suggest that V1/V2 from a given gp120 protomer may be in close proximity to the CoRbs on a neighboring protomer to mask the coreceptor site (44).

2. Glycan shield: HIV can also use glycans to occlude antibody epitopes on gp120; 50% of the molecule is covered with carbohydrates that render the underlying protein surface invisible to the immune system. The virus can also shift the locations of its

glycans in vivo (45–47). These observations have led to the proposition of a dynamic or so-called evolving glycan-shield model (46) by which, through the continuous repositioning of its glycans, HIV is able to escape type-specific neutralizing antibody responses. The repositioning of glycans may also affect local protein folding and, hence, also indirectly affect neutralizing antibody binding.

3. Entropic masking: It has been postulated that the productive interaction of antibodies with conserved regions on the HIV-1 envelope spike, in particular with the receptor binding sites, may be limited as a result of intrinsic entropic barriers that are imposed upon antibodies within the context of the viral spike (48). Large changes in entropy are generally observed upon binding of nonneutralizing or weakly neutralizing antibodies to receptor sites on monomeric gp120, whereas small changes in binding entropy are typically observed with broadly neutralizing antibodies (49, 48). Neutralizing antibodies, in contrast, do not incur such an entropic penalty because they require relatively less conformational reorganization within the gp120 molecule for efficient binding. Functional oligomeric gp120 is likely more conformationally fixed owing to interaction with neighboring gp120 protomers within the spike (50), which supports the notion of entropic masking.

4. Substitution of nonessential residues: antibody recognition of the CD4bs may be easily circumvented by the virus if the ensemble of residues that are critical for antibody binding do not closely match those that are essential for CD4 binding, even if the epitope maps of both ligands overlap geometrically (51, 52).

Strategies for designing HIV immunogens
At present, many antigen design strategies are being pursued to evaluate their potential to elicit broadly neutralizing antibodies.

1. Mimicry of functional HIV Envelope spike: An immense hurdle in the design of antigens that mimic the envelope spike is the stabilization of the gp120-gp41 interaction, which is normally mediated by noncovalent interactions and is

relatively labile. An often-used stabilization strategy is mutation of the protease cleavage site between gp120 and gp41 so that the two subunits remain covalently linked upon protein expression (53–61). However, uncleaved envelope proteins appear to be antigenically distinct from their cleaved counterparts (62). Thus, whereas both non neutralizing and neutralizing antibodies bind uncleaved envelope glycoproteins, only neutralizing anti-gp120 antibodies efficiently bind oligomeric glycoprotein when the gp160 precursor protein is effectively cleaved by furin proteases (or equivalent proteases). As an alternative approach to this problem, single cysteine residues have been introduced into gp120 and gp41, which results in the formation of an intermolecular disulfide bridge upon expression of the proteolytically cleaved oligomer (62, 63, 64). However, the disulfide bridge stabilized complexes still display epitopes that are recognized by nonneutralizing antibodies (134), which suggests that mimicry of the native envelope glycoprotein is not fully achieved with these molecules. A further challenge in designing envelope spike mimics is the preservation of the trimeric state of the antigen complex. Thus, although mutation of the protease cleavage site or the introduction of a disulfide bridge allows gp120 and gp41 to remain covalently linked, soluble oligomers derived in this manner often either disassociate into single heterodimers of gp120 and gp41, or tend to assemble into varying oligomerization states, such as dimers, trimers and/or tetramers (54, 57–59, 61, 65, 66). One of the strategies that have been employed to circumvent these problems is covalent linkage of heterologous trimerization domains to the C terminus of recombinant oligomeric proteins (67–69). These domains indeed improve trimerization, and the corresponding proteins elicit neutralizing antibodies at modestly improved levels compared with preparations that contain a mixture of oligomeric proteins (70). Notably, immunization studies show that purified trimeric glycoproteins—independent of whether or not they are stabilized by a heterologous trimerization domain—are somewhat superior at eliciting cross-neutralizing antibodies as compared to simple monomeric gp120 formulations (71–73).

2. Other antigen design strategies have attempted to present the oligomeric envelope spike in a more native environment, for example, through mild chemical inactivation of virus particles (which appears to preserve the oligomerization state and functionality of the envelope spike) (74, 75) or, alternatively, through the incorporation of uncleaved oligomeric envelope glycoprotein complexes into proteoliposome preparations (76). Though promising, both of these two approaches have so far failed to elicit neutralizing antibodies of the desired breadth and potency (71, 77, 78).

3. Another attractive approach to presenting the envelope spike in a native-like context is the generation of virus-like particles (VLPs) (79). VLPs represent a unique class of subunit antigens, as they may allow closer mimicry of structures on the surface of virions compared with other types of antigen presentation. VLPs also have the advantage that, in contrast to soluble subunit antigens, they are able to stimulate both cellular and humoral responses, although, unlike live or inactivated virus, they do not contain viral genetic material (79). So far, however, the VLP approach has not achieved the desired success in eliciting high levels of cross-neutralizing antibodies (80–82). A possible reason for the inability of current VLP preparations to elicit more broadly neutralizing antibodies may be due to the expression of both nonfunctional as well as functional forms of the spike on the surface of the virus particle. The elicitation of antibodies to cellular proteins that are incorporated into the virus particle may also limit the efficacy of VLP preparations.

4. To enhance the elicitation of broadly neutralizing antibodies, variant trimeric envelope glycoproteins have been designed with truncated or altered variable regions. In some cases, alterations have also been made in gp41. Although some of these alterations have resulted in an improvement in the neutralizing capacity of the immune sera, additional modifications and other improvements will be necessary to significantly boost the ability of these envelope spike mimics to elicit antibodies that are capable of neutralizing a wide range of primary viral isolates.

5. Epitope focused immunogen design: Epitope-focused antigen design may provide a means by which to limit the induction of non-neutralizing antibodies, while preserving the elicitation of antibodies to more conserved epitopes.

a. CD4bs immunogens: The molecular interactions of broadly neutralizing antibodies to the CD4bs have suggested at least four strategies to elicit these antibodies (fig).

First, trimeric forms of HIV-1 Env have been generated by inclusion of the gp41 trimerization sites after deletion of the transmembrane domain. The trimeric protein can be further stabilized by addition of trimerization sequences from such proteins as the fibritin protein from phage lamda. Such trimers can be further stabilized with site-specific mutations previously shown to fix the core structure (82). In these prototypes, the variable V1-V3 domains are often removed to minimize immune responses to irrelevant strain-specific structures.

An alternative approach is to develop immunogens based on a monomer structure (Fig. second panel). Such proteins have been derived from stabilized core Env proteins further altered based on an understanding of structure (83;84). The surface of the conformationally stabilized Env core protein was modified and masked further with glycans. Such probes have been used to analyze antisera for the presence of broadly neutralizing antibodies and served also as prototype immunogens to elicit antibodies to this site.

A third approach focuses on generating a subdomain of the HIV-1 Env, the outer domain that contains the initial site of CD4 attachment (Fig, third panel). In this protein, a large portion of the inner domain that elicits nonneutralizing antibodies is eliminated. The immune response is therefore directed to the relevant site of initial CD4 binding. In addition, by removing parts of the inner domain required for CD4 binding, potential inhibition by CD4 attachment is avoided. In theory, modified forms of the outer domain will allow targeting of the immune response to the most relevant conserved region of binding. Mutations in the outer domain region have

been designed to preserve high-affinity binding and are currently under evaluation both as probes of serum neutralizing antibody activity and as immunogens.

A fourth approach to CD4bs immunogen development employs scaffolds based on informatics and epitope transplantion (Fig, lower panel) (e.g., adding the b15 loop to an unrelated structure that presents the epitope naturally). Scaffolds have been identified that bind CD4bs antibodies and are the subject of continued investigation.
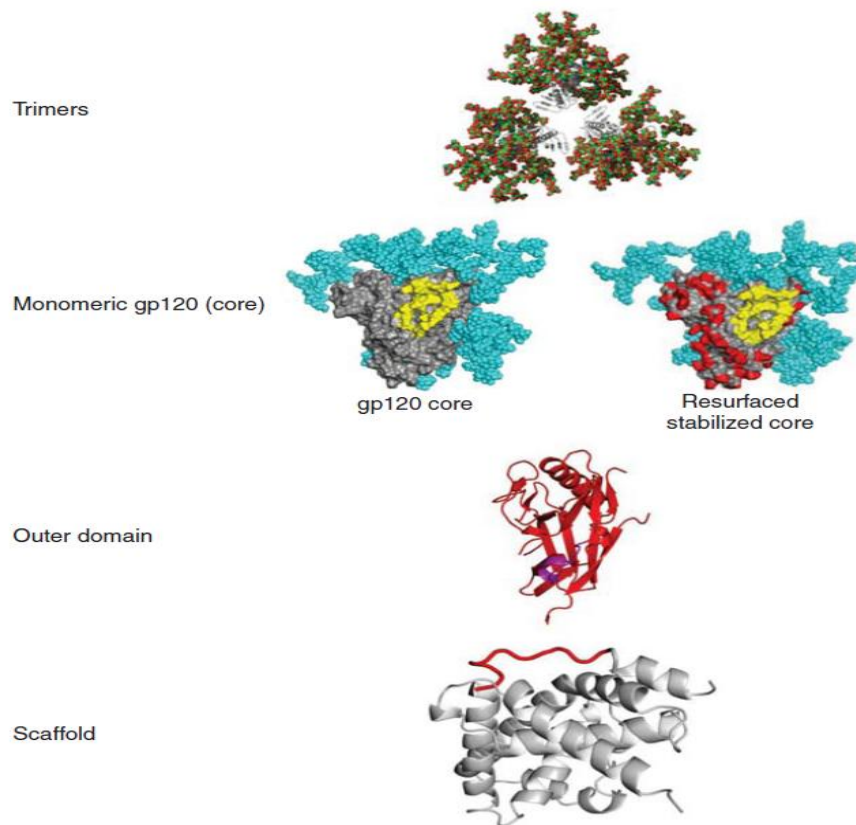
Figure 4: Alternative forms of HIV Env serve as prototype immunogens for neutralizing antibody vaccines. Different forms of the HIV Env can be used to elicit neutralizing antibody responses. They range from the most complex form, the HIV trimer that most closely resembles the form found on the viral spike (upper row). monomeric forms, which include the gp 120 core or resurfaced stabilized cores (second row), a region of the core that is composed primarily of the outer domain (0D) which includes the CD4-binding site (third row), or selected subdomains, such as the CD4-binding b-15 loop or M PER attached to a heterologous stable scaffold (bottom row). Modification of these prototypes by deletion of variable regions, removal, or addition of glycans, stabilization with disulfide bonds or addition of space filling mutations can serve to alter immunogenicity and elicit antibodies of the desired specificity.

b. Glycan and quaternary immunogens: At least two types of antibodies recognize carbohydrate determinants on HIV-1 Env. Among these antibodies are those like the prototypic 2G12 mAb that recognize high-mannose structures on the outer domain.

There is concern that it will be difficult to generate antibodies to carbohydrate structures added by endogenous glycosylation machinery, and even if it is possible, there is concern that such antibodies may react with such carbohydrates on host proteins. Nonetheless, several groups have attempted to develop immunogens by chemical conjugation with nonhuman glycans or by selection in yeast (85). Although some structures have been defined that react with 2G12, it has not yet been possible to elicit such neutralizing antibodies with these immunogens. Another class of exceptionally neutralizing antibodies has been directed to glycans, and possibly peptide sequences in the V1/V2 region, with additional interactions dependent on the V3 region. Such immunoglobulins have been termed quaternary antibodies, although it is not certain that they are directed to complex conformational determinants from different parts of Env. Although the structure of at least one such antibody has been elucidated, the molecular details of its interaction with Env remain hypothetical. Efforts at immunogen design have focused thus far on membrane-bound trimers, which show the strongest binding to these antibodies; however, occasional monomeric gp120 derivatives have been identified and are also under investigation. Although this class of antibody represents _25% of the antibodies in the sera from subjects with broadly neutralizing responses (86), it has not yet been possible to elicit these antibodies by vaccination in animal models or humans.

c. MPER, V3 and CD4i Immunogens: At least two broadly neutralizing mAbs, 2F5 and 4E10, recognize the highly conserved MPER region of Env. Both show reasonable breadth of neutralization, although their potency is generally low. Structures of these antibodies complexed to their cognate peptides have been determined and suggest that hydrophobic patches are required for stable interactions needed for neutralization. In the case of 4E10, efforts to develop immunogens based on stabilized peptides have allowed the definition of vaccine candidates with appropriate antigenic profiles but they do not elicit neutralizing antibodies. In the case of 2F5, a variety of approaches have elicited antibodies that react with 2F5 peptides but these antibodies do not neutralize diverse viral strains. An

understanding of the structural interactions of 2F5 with its viral target has suggested that the structure of the 2F5 peptide in the context of gp120 as well as its interaction with the viral membrane through a hydrophobic patch is required for neutralization. Based on these findings, Ofek and colleagues have generated scaffolds that present the constrained 2F5 peptide which can elicit antibodies that interact similar to the 2F5 ab, suggesting a potential strategy to elicit such neutralizing antibodies (87).

The V3 and CD4i regions have also received considerable attention with respect to immunogen design and have been reviewed elsewhere. It is possible to elicit neutralizing antibodies to these V3 epitopes using modified murine retroviral gp or HIV-1 Env proteins. Because the V3 region is not exposed on most naturally circulating viral isolates, it is unlikely that these responses would be effective in the absence of an activity that would increase its exposure. Thus far, no such agonistic antibodies have been identified and this target therefore poses a considerable challenge. Similarly, CD4i antibodies can also be readily induced after immunization with HIV Env stabilized core proteins (88). Although such antibodies are seen frequently in HIV-infected individuals, they do not mediate neutralization although it remains possible that they may contribute to protection through ADCC function.

Epitopes and epitopic vaccines

An epitope can be simply defined as that part of an antigen involved in its recognition by an antibody. It is the minimal structure necessary to invoke an immune response. In some senses, it is the immunological quantum that lies at the heart of immunity. Epitopes must come from proteins accessible to immune surveillance and to be specific for a particular pathogen or tumour.

T cell epitopes are short peptides bound by major histocompatibility complexes (MHC) and subsequently recognized by T cells. B cell epitopes are regions of the surface of a protein, or other biomacromolecule, recognized by soluble or membrane-bound antibody molecules. It is the recognition of epitopes by T cells, B cells and soluble antibodies that lies at the heart of the immune response which, in turn, leads to activation of the cellular and humoral

immune systems and, ultimately, to the effective destruction of pathogenic organisms. The accurate prediction of B cell and T cell epitopes is thus the pivotal challenge for immunoinformatics. Although the importance of nonpeptide epitopes, such as carbohydrates and lipids, is now increasingly well understood, peptidic B cell and T cell epitopes remain the principal goal [89]

Vaccines can be based either on epitopes or on sets of epitope: the epitope is the minimal structure necessary to invoke an immune response. In some senses, it is the immunological quantum that lies at the heart of immunity. Epitopes must come from proteins accessible to immune surveillance and to be specific for a particular pathogen or tumour. Targeting amino acid sequences as epitopes, as recognized by neutralizing antibodies or which bind MHCs, rather than whole protein antigens has the advantage that many sequences able to induce autoimmunity are eliminated. Likewise, these vaccines are potentially much safer: they contain no viable self-replicating microorganisms and cannot cause microbial disease [89].

Any peptide-based vaccine will need adjuvant to enhance immunogenicity. Whole organisms contain not one or even a handful of epitopes, but vast numbers of all kinds, as well as abounding with adjuvant-like natural products and danger signals. Thus it is little surprise that single epitopes are not as immunogenic as, say, live attenuated vaccines. The ability to tailor epitopes in light of our understanding of MHC restriction suggests that epitope based vaccines will play an important role, particularly in cancer vaccines [89]

Epitope prediction

Identification of epitopes that invoke strong responses from B-cells is one of the key steps in designing effective vaccines against pathogens. Because experimental determination of epitopes is expensive in terms of cost, time and effort involved, there is an urgent need for computational methods for reliable identification of B-cell epitopes.

Although several computational tools for predicting B-cell epitopes have become available in recent years, the predictive performance of existing tools remains far from ideal.

For a vaccine to be effective it must invoke a strong response from both T Cells and B Cells; therefore, epitope mapping is a central issue in their design. *In silico* prediction methods can accelerate epitope discovery greatly. B Cell and T Cell epitope mapping has led to the predictive scanning of pathogen genomes for potential epitopes. There are over 4,000 proteins in the TB genome; this means that experimental analysis of host-pathogen interactions would be prohibitive in terms of time, labour, and expense.

Bcell epitope prediction is more problematic due to the difficulties in correctly defining both linear and discontinuous epitopes from the rest of the protein. The epitope of a B Cell is defined by the discrete surface region of an antigenic protein bound by the variable domain of an antibody. The production of specific antibodies for an infection can boost host immunity in the case of both intracellular and extracellular pathogens. The antibody's binding region is composed of three hypervariable loops that can vary in both length and sequence so that the antibodies generated by an individual cell present a unique interface. All antibodies contain two antigen-binding sites, composed of complementary determining (CDR) loops. The three CDR loops of the heavy and light chains form the 'paratope', the protein surface which binds to the antigen. The molecular surface that makes specific contact with the residues of the paratope is termed an 'epitope'. A Bcell epitope can be an entire molecule or a region of a larger structure. The study of the paratope–epitope interaction is a crucial part of immunochemistry, a branch of chemistry that involves the study of the reactions and components on the immune system.

Despite the extreme variability of the region, the antibody-binding site is more hydrophobic than most protein surfaces with a significant predilection for tyrosine residues.

BCell epitopes can be divided into continuous (linear) and discontinuous (conformational), the latter being regions of the antigen separated within the sequence but brought together in the folded protein to form a three-dimensional interface.

Continuous epitopes correspond to short peptide fragments of a few amino acid residues that can be shown to cross-react with antibodies raised against the intact protein. Since the residues involved in antibody binding represent a continuous segment of the primary

sequence of the protein they are also referred to as 'linear' or 'sequential' epitopes. Studies have shown that this class of epitope often contains residues that are not implicated in antibody interaction, while some residues play a more important role than others in antibody binding.

Discontinuous epitopes are composed of amino acid residues that are not sequential in the primary sequence of a protein antigen but brought into spatial proximity by the three dimensional folding of the peptide chain (91).

There is considerable interest in developing reliable methods for predicting B cell epitopes. However, to date, the amino acid distribution of the complementary antigen surface has been difficult to characterize, presenting no unique sequential or structural features upon which to base a predictive system. It is partly for this reason that B cell epitope has lagged far behind T Cell prediction in terms of accuracy but also because most of the data upon which predictions are based remain open to question due to the poorly understood recognition properties of cross reactive antibodies. One of the central problems with B cell epitope prediction is that the epitopes themselves are entirely context dependent. The surface of a protein is, by definition, a continuous landscape of potential epitopes that is without borders. Therefore both epitope and paratope are fuzzy recognition sites, forming not a single arrangement of specific amino acids but a series of alternative conformations.

In this instance, a binary classification of binder and non-binder may simply not reflect the nature of the interaction. A factor also to be considered is that the average paratope consists of only a third of the residues within the CDR loops, suggesting the remaining two-thirds could potentially bind to an antigen with an entirely different protein surface. Often a short length of amino acids can be classified as a continuous epitope, though in fact it may be a component of a larger discontinuous epitope; this can be a result of the peptide representing a sufficient proportion of the discontinuous epitope to enable cross-reaction with the antibody. Since the majority of antibodies raised against complete proteins do not cross-react with peptide fragments derived from the same protein it is thought that the majority of epitopes are discontinuous.

It is estimated that approximately 10% of epitopes on a globular protein antigen are truly continuous in nature.

Several experimental techniques are currently available for experimental mapping of B-cell epitopes. However, the high cost and effort involved makes them impractical for application on a genomic scale. Computational techniques offer a fast, scalable, and cost-effective approach for predicting B-cell epitopes, for focusing experimental investigations and for improving our understanding of antigen-antibody interactions. Hence, there is a growing interest in the development of sophisticated computational tools for reliable prediction of B-cell epitopes.

Recent studies have pointed out some of the limitations of current epitope prediction methods [91]. Hence, increasing the reliability of computational methods for B-cell epitope prediction remains a major challenge in computational vaccinology.

Predicting linear BCell epitopes

In the case of linear B-cell epitopes, antibody-antigen interactions are often conformation-dependent. The conformation-dependent nature of antigen-antibody binding complicates the problem of B-cell epitope prediction. Hence, B-cell epitope prediction is less tractable than T-cell epitope prediction.

Major approaches for predicting linear Bcell epitopes have been as follows:

Propensity scale methods: The first propensity scale method for predicting linear B-cell epitopes was introduced by Hopp and Woods [27] and utilized Levitt hydorophilicity scale [28] to assign a propensity value to each amino acid. This method is based on the assumption that antigenic determinants of protein sequences correspond typically to sequence windows that contain a large number of charged and polar residues and lack large hydrophilic residues. Subsequently, several other propensity scales have been proposed for predicting linear B-cell epitopes. For example, Parker et al. [23], Karplus et al. [24], Pellequer et al. [29] and Emini et al. [26] have proposed propensity scale based methods that use hydrophilicity, flexibility, turns, or solvent accessibility propensity scales (respectively). PREDITOP [25], PEOPLE [30], BEPITOPE [31], and BcePred [32] predict

linear B-cell epitopes based on combinations of physico-chemical properties as opposed to propensity measures that rely on individual properties.

Machine learning methods: several authors have explored machine learning based methods for predicting linear B-cell epitopes using amino acid sequence information. B-cell epitopes using amino acid sequence information. ABCPred [12] uses recurrent artificial neural networks for predicting linear B-cell epitopes and was evaluated on a dataset of 700 B-cell epitopes and 700 non-epitope peptides using 5-fold cross validation tests.

BCPred [5] and FBCPred [10] predicts linear B-cell epitopes and flexible length linear B-cell epitopes (respectively) using support vector machine (SVM) classifiers that use string kernels [33]. COBEpro [34] uses a two-step procedure for predicting linear B-cell epitopes. In the first step, an SVM classifier is used to assign scores to fragments of the query antigen. The input of the SVM is a vector of similarities between the input fragment and all training peptide fragments. In the second step, a prediction score is associated with each residue in the query antigen based on the SVM scores for the peptide fragments.

Using several benchmark datasets, COBEpro has been shown to achieve a competitive performance with other linear B-cell epitope prediction methods.

Predicting conformational Bcell epitopes

Although more than 90% of B-cell epitopes are estimated to be conformational in nature, most experimental as well as computational methods focus on mapping linear B-cell epitopes. However, in the past few years, there is increasing interest in methods for predicting conformational B-cell epitopes.

It has been estimated that >90% of B-cell epitopes are discontinuous, i.e., consist of segments that are distantly separated in the pathogen protein sequence and brought into proximity by the folding of the protein. Identification of discontinuous epitopes is difficult, since the complete analysis must be done in context of the native antigen structure. The most informative and accurate method for identification of discontinuous epitopes is determination of structures of antigen–antibody complexes by X-ray crystallography. The use of discontinuous epitopes derived from presently available X-ray structures is

complicated by two major problems: First, the available data on discontinuous epitopes in different antigens is much reduced compared to linear epitopes; second, very few antigens have been studied to completely identify various discontinuous epitopes in the same antigen. The existence of undetected epitopes that are not identified in the data set can make it harder to develop good prediction algorithms because they influence the measured performance. However, detailed structural knowledge on antibody–antigen complexes is growing, and allows for broader analysis of discontinuous epitopes in various antigens and development of better prediction methods.

Three major approaches for predicting conformational Bcell epitopes are:

Sequence based prediction methods: Sequence-based methods in predicting conformational B-cell epitopes have the advantage that they do not require the structure of the target antigen to be known. The amino acid propensity scale methods that assign a prediction score to each residue in the antigen sequence can in principle be used to predict conformational B-cell epitopes. Such methods provide a baseline for evaluation of more sophisticated conformational B-cell epitope prediction methods. A large body of work using machine learning methods for predicting protein-protein [35,36], protein-DNA [37,38], and protein-RNA [39,40] interfaces using sequence-derived features has demonstrated the feasibility of using sequence-based classifiers in reliably identifying functionally important sites in proteins. The development of such B-cell epitope predictors would make it feasible to identify conformational B-cell epitopes in antigenic sequences for which no solved 3D structures are available.

Structure based prediction methods: The most accurate experimental method for identifying conformational B-cell epitopes relies on determination of the structure of antigen-antibody complexes using X-ray crystallography. Because the number of solved antigen-antibody complexes, or for that matter, the solved antigen structures, is small relative to the number of available antigenic sequences, there are only a small number of methods that utilize 3D structure derived information in predicting conformational B-cell epitopes.

One of the first conformational B-cell epitope predictors is the conformational epitope predictor (CEP) [7]. Given an antigen with a known structure, CEP uses accessibility of residues and spatial distance cut-off to predict linear and conformational B-cell epitopes.

DiscoTope, a method developed by Andersen et al. [9], uses a combination of amino acid statistics, spatial context, and surface accessibility of amino acids to predict conformational B-cell epitopes. DiscoTope has been shown to outperform propensity scale methods on a dataset of 76 antigen-antibody complexes. The same study also showed that predictors that combine both sequence and structure-derived features of antigens are more accurate than those that rely on either sequence or structure derived features alone.

The B-cell conformational epitope predictor proposed by Rapberger et al. works as follows (given the 3D structure of a query antigen):

Fast atomic density evaluation (FADE) [44] is applied to select an antibody among a library of 26 available antibodies showing best shape complementarity to the target antigen;

FastContact algorithm [45] is used to identify the most likely interaction site between the selected antibody and the target antigen; (iii) Antigen residues that show a decrease in relative solvent accessible surface area (estimated using a probe size of 3Å) of at least 20% in the complex are predicted as belonging to a discontinuous epitope. This method was shown to outperform the CEP method using a non-redundant dataset of 26 antigen-antibody complexes from Protein Data Bank (PDB).

Mimotope analysis based prediction methods: In this approach, a phage display library of random peptides is scanned against an antibody of interest to obtain a panel of peptides (called mimotopes) that bind to the antibody with high affinity. It is assumed that this panel of mimotopes mimics the physico-chemical properties and spatial organization of the genuine epitopes. However, the precise identification of the epitope mimicked by the set of mimotopes is not straightforward since the epitope is often discontinuous (conformational) and the epitope and mimotopes do not necessarily share a high degree of sequence similarity. Moreover, some of the mimotopes may reflect noisy biological observations and should be filtered out in the analysis. Hence, several computational

26

methods for localizing the panel of affinity selected peptides on the surface of a target antigen have been proposed in literature [60-65].

In general, mimotope analysis methods available in the literature differ from each other in terms of:

how they represent the antigen structure/sequence;

how they align mimotopes with the target antigen structure/sequence;

how they cluster the mimotopes and rank the predicted epitopes.

Epitope mapping: Although physicochemical and structural features do not seem to predict the location of B-cell epitopes, it can be anticipated that combination of these features with specific amino acid sequences or motives may constitute the leap forward to accurate tools for predicting B-cell epitopes. Batori et al. [4] were the first to provide evidence suggesting the existence of general antibody-binding motifs. The Epitope Mapping Tool (EMT) seems to be the only tool providing a concrete epitope amino acid sequence useful for peptide-based vaccine development. A similar approach is provided by the search algorithm Pep-3D-Search. This algorithm should facilitate the identification of peptide sequences and motifs obtained by phage display by creating a surface graph using all surface residues of the protein. On a set of eight epitopes and two protein-binding sites defined by crystallography, the Pep-3D-Search revealed a sensitivity and PPV of 36% and 70%, respectively [32 from1-s2.0-51740674909]

PROTEIN ENGINEERING

Epitope backbone versus side chain grafting

Transplantation of functional motifs onto heterologous proteins is a general approach to engineer novel molecules, such as antigens, protein inhibitors, or enzymes. The majority of previous protein grafting studies relied on grafting the side chains of a given functional motif onto regions of high structural homology of candidate scaffold proteins. These approaches are restricted by the existence of structurally characterized scaffolds with exposed backbone similar in conformation to a particular motif. To avoid this limitation, numerous efforts involved the grafting of motifs located on common secondary-structure

elements, such as single α-helices and polyproline type II helices, α-helical coiled coils and β-sheets to analogous regions on heterologous proteins. In contrast, backbone grafting of linear motifs requires reduced structural similarity between a functional motif and candidate scaffolds, since only the termini of the motif, rather than the entire backbone length, have to be structurally matched with a scaffold for transplantation.

Furthermore, backbone grafting has the added ability to alter the backbone conformation of a candidate scaffold and to engineer interactions between the transplanted motif and the scaffold such that the desired structural conformation of the motif is stabilized. Therefore, backbone grafting may allow transplantation of structurally diverse linear functional motifs when no suitable scaffolds exist for side-chain grafting. (119,120,121)

Epitope backbone grafting is a computational protocol that

(A) Identifies scaffolds that can accommodate the backbone conformation of a given linear epitope.

(B) replaces corresponding native scaffold backbone with the desired backbone conformation of an epitope

(C) Integrates the epitope with the native scaffold such that the desired epitope conformation is stabilized, and

(D) Ensures that the antibody can interact with the resulting epitope scaffold.

The computational procedure is divided into two stages—matching and design.

MATCHING: the matching procedure aims to search suitable sites on the candidate scaffolds that can be replaced for backbone grafting of the epitopes of interest (continuous or discontinuous). In case of grafting of discontinuous epitopes, the matching algorithm makes sure that all the discontinuous stretches are matched to a particular candidate scaffold. Single matches are not reported in case of discontinuous epitopes; rather multiple matches are searched and reported. Matching makes use of different alignment systems namely N2C, C2N, S and E to determine the matches.

In N2C alignment, the backbone atoms (N, C, and Cα) of the N terminus residue of the epitope is aligned on candidate scaffolds and RMSD values are measured between the C terminus of the epitope and proximal residues on the scaffold.

In C2N alignment, the backbone atoms (N, C, and Cα) of the C terminus residue of the epitope is aligned on candidate scaffolds and RMSD values are measured between the N terminus of the epitope and proximal residues on the scaffold.

In S alignment, backbone superimposition is performed.

In E alignment, the three backbone atoms of the epitope loop termini are superimposed onto the corresponding atoms of the residue pairs on the query scaffold.

Scaffolds with rmsd values within a certain threshold set by the user and with minimum clashes are considered for further designing. Two types of Rosetta clashes are checked for in order to determine biologically significant scaffolds namely:

Intra clashes: these are the clashes between the epitope (single motif or continuous/ multi-segment motif or discontinuous) and the scaffold. Intra clash values have to be as low as possible so as ensure that the matches made do not destabilize the scaffold structure and are on the surface.

Inter clashes: these are the clashes at the interface and these too have to least so that the designed scaffolds do not bump into the antibody structure. Higher values of inter clashes also indicate that matches made may not be on the surface.

So the filters and checks for finding suitable and biologically significant matches includes the threshold rmsd values and inter and intra clash threshold values set by the user.

DESIGN: In the design phase, the selected scaffolds after passing the filters are used for grafting the desired functional motifs. This includes replacing the matched segments of the scaffold with the desired motifs and closing the gaps between them by adding linkers of suitable length. Hence designing essentially covers both the sequence as well as structure design.

BROADLY NEUTRALZING ANTIBODIES

Broadly neutralizing antibodies are key players in mediating protection against HIV by targeting certain conserved regions on gp120 surface protein of the virus and hence effectively neutralizing many types of the HIV. With the discovery of these broadly neutralizing antibodies, new approaches to HIV vaccine development have emerged that exploits the potential of designing immunogens carrying the conserved regions (of gp120) targeted by these antibodies. These immunogens if administered as a vaccine may elicit such broadly neutralizing antibodies within the vaccinated individuals. Studies in the field of AIDS vaccine research and the technological advances have now made it possible to isolate and characterize these antibodies even though they form a minor portion of the total antibody repertoire in chronically infected HIV patients. Also the exact positioning of these antibodies on the Env protein has also been well understood. So far, these antibodies have been known to target the CD4 binding site, membrane proximal external regions MPER, conserved regions of the variable loop V1/V2/V3 of HIV gp120 and the glycan dependent epitopes (122)These antibodies display certain unusual characteristics like polyreactivity, extensive somatic hypermutations and long VH CDR3 which not only make them very effective against HIV infections but also very important for developing new strategies for prevention against AIDS (123).
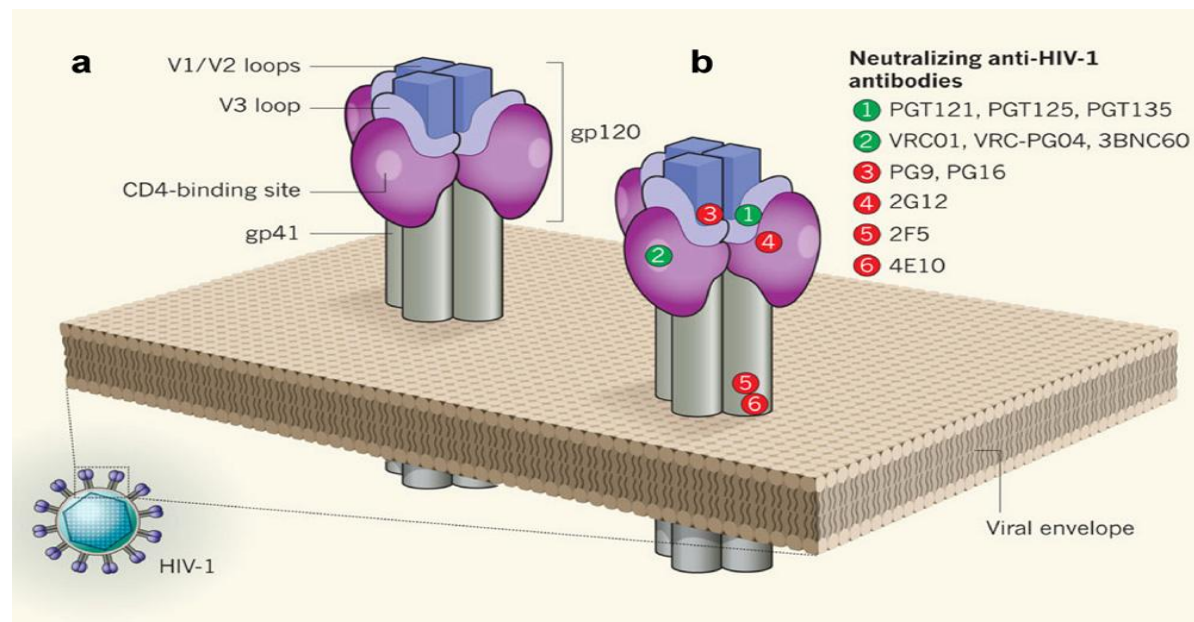
Figure 5: The envelope of HIV-1 carries spikes. (a) Each spike is made of three molecules of the surface glycoprotein gp120 and three molecules of the transmembrane glycoprotein gp41. Glycoprotein gp120 contains variable V1/V2 and V3 loops, as well as the binding site for CD4. (b) The binding sites of broadly acting and potent HIV-1- specific neutralizing antibodies are shown as colored circles

.

For many years, the HIV antibody field was restricted to the study of a few monoclonal antibodies that identify the following three epitope sites on the envelope: the CD4-binding site, the membrane proximal external regions, and a particular pattern of glycans.

These antibodies are very rare, out of thousands made that are less active, and most of the protective antibodies have unusual features, such as a long complementarity-determining region 3 (CDR3) in the heavy-chain variable (VH) region, a large number of somatic mutations and polyreactivity16.

It is now clear that ~20% of patients chronically infected with HIV-1 make BnAbs, although it takes them 3–4 years to do so(124). It is now possible to isolate and characterize these antibodies, even though they form a tiny minority of the total antibody response to Env in any patient (125-127). Thus, through the cloning of antigen-specific Bcells or by PCR amplification of genes encoding antibodies in plasmablasts obtained from infected patients, it has been possible to characterize hundreds of new monoclonal antibodies that are broadly cross-reactive in their specificities.

31

In many cases, the exact positioning of the antibody on the Env protein is understood, which explains often why these antibodies are neutralizing, whereas others that bind nearby are not. Furthermore, a published report of an HIV-1- infected person who made both BnAbs to the CD4-binding site and V2V3-conformational BnAbs has established proof of the concept that it is possible to induce multiple types of BnAbs in the same person.

BnAbs are unusual antibodies in that they are transcribed from genes with considerable somatic mutation, often containing more than 50 somatic mutations relative to the sequence of their nearest germline ancestor. They are often polyreactive with non-HIV antigens, and for some specificities, such as the V2V3 conformational epitope, they frequently have extremely long VH CDR3 regions. When the germline-encoded ancestors of the BnAbs to HIV-1 Env are derived, they often react only with low or undetectable affinity with the HIV-1 envelope. Immune-tolerance mechanisms are responsible for much of the control of the production of antibodies with the BnAb traits of polyreactivity, long VH CDR3 regions and extensive somatic hypermutations16. That last trait might explain why it takes a long time (2–3 years) for such antibodies to appear and why they are more prevalent in patients with high virus loads, in whom there might be a greater antigenic stimulus (124).

A recent review (128) reports that NAbs tend to increase in potency over time and broadly cross neutralizing (BCN) responses, capable of recognizing heterologous HIV-1 variants, develop in a subset of individuals after primary infection. The reasons why some individuals develop BCN antibodies is unclear but is related to the duration of infection and viral levels suggesting that years of persistent viral stimulation are necessary for their generation. However, the finding that not all individuals develop BCN antibodies points to the role of additional factors that could be viral or host or most likely interplay between the two. There is some evidence that viral genetic subtype may also be important as the breadth and potency of humoral responses is reported to be higher in subtype C and A than in subtype B infections (129). A large number of studies published in the last few years have described the presence of BCN antibodies in different cohorts. Although there is no standard definition of breadth, all studies identified a proportion (_30%) of individuals with BCN antibodies based on their ability to neutralize a significant number of heterologous viruses usually of multiple genetic subtypes. In some cases, the specificities of

the antibodies conferring breadth have been mapped and are reactive with conserved envelope regions, as exemplified by well-characterized neutralizing MAbs IgG1b12 (anti-CD4bs), 4E10 (carboxy-terminal MPER), 2F5 (amino-terminal MPER), and 2G12 (glycan array on gp120 outer domain). As new NAb epitopes are discovered, mapping tools are expanding to detect these specificities. Thus, PG9/16-like antibodies are frequently being identified in BCN sera (Fig.). However, not all the BCN activities have been accounted for, suggesting that additional important neutralizing epitopes on the HIV-1 envelope glycoprotein are yet to be discovered.

Intensive efforts are underway by many laboratories to isolate MAbs from these individuals as this may lead to the identification of new sites on the HIV envelope that could serve as targets for vaccine induced immune responses.

It has been proposed that BnAbs may naturally arise only in subjects with relatively relaxed immune tolerance controls and that one way to induce such antibodies by a vaccine is to design artificial immunogens that specifically stimulate ancestral and early-intermediate antibodies of a BnAb clonal lineage (130).

The question of whether breadth is conferred by single or multiple specificities has in part been addressed by some of these studies. Certainly, cases of multiple specificities have been documented, but there is also evidence that single MAbs can account for the breadth displayed by BCN plasma, most recently through the isolation of VRC01, a highly potent and broad anti-CD4bs mAb from an individual whose plasma neutralizing activity was shown to have anti-CD4bs activity. On the other hand, a study by Nussensweig and colleagues suggested that neutralization breadth is the result of multiple antibody specificities, each one targeting a few viral variants. For vaccine design, it would likely be considerably more challenging to induce multiple rather than single specificities because different antigens and thus a more complex vaccine would be required to stimulate a polyclonal response. (128)

Another important question is whether escape from BCN antibodies occurs as readily as it does from autologous NAbs. There is less data on this but one might imagine that because BCN antibodies target more conserved sites, escape would be more difficult to achieve.

The development of neutralization breadth may include a requirement for extensive somatic mutational diversification that may be impeded by the B cell dysfunction that occurs in HIV-1 infection. A role for affinity maturation in the ontogeny of anti-HIV-1 neutralizing capacity is supported by Toran and coworkers, who showed that all anti-gp120 Nabs in a long-term nonprogressor were clonally related with considerable somatic hypermutation.

Furthermore, immunogenetic analysis of existing MAbs suggests that they have undergone multiple rounds of affinity maturation to achieve BCN activity. These MAbs are characterized by unusual physical features such as heavy chain domain-swapping, sulphated tyrosines, long and hydrophobic CDRH3's, and high levels of hypermutation (131). Whether these events are required for antibodies to mediate cross-reactivity is not known, and longitudinal studies of immunoglobulin genes among those individuals who develop breadth are needed to understand how the immune system can generate such antibodies. Vaccinations strategies will need to be revised to not only consider the immunogen used, but to incorporate the required number of booster and/or antigen exposures needed to reach the required level of antibody affinity.

VRCO1
The heavy chain of VRC01 interacts with gp120 in a manner similar to CD4. A 43° rotation coupled with a 6-Å shift from the CD4-defined orientation focuses VRC01 onto the conformationally invariant site of initial CD4 attachment, allowing it to overcome the masking that diminishes the neutralization potency of most CD4-binding-site antibodies. To achieve this mode of recognition, VRC01 contacts gp120 mainly through V-gene derived regions substantially altered from their genomic precursors. Partial receptor mimicry and extensive affinity maturation thus facilitate effective neutralization of HIV-1 by natural human antibodies.

VRCO1 is not only broadly neutralizing but is also present in high titers in humans [131]. While potentially accessible, the CD4-binding site is protected from humoral recognition by glycan and conformational masking (132).

The interactive surface between VRC01 and gp120 encompasses almost 2500 Å2, 1244 Å2 contributed by VRC01 and 1249 Å2 by gp120 (131). On VRC01, both heavy chain (894 Å2) and light chain (351 Å2) contribute to the contact surface, with the central focus of binding on the heavy chain second complementarity-determining region (CDR H2). Over half of the interactive surface of VRC01 (644 Å2) involves CDR H2, a mode of binding reminiscent of the interaction between gp120 and the CD4 receptor; CD4 is a member of the V-domain class of the immunoglobulin superfamily (133), and the CDR2-like region of CD4 is a central focus of gp120 binding (134). For CD4, the CDR2-like region forms antiparallel, intermolecular hydrogen bonds with residues 365-368gp120 of the CD4-binding loop of gp120 (134); with VRC01, one hydrogen-bond is observed between the carbonyl of Gly54VRC01 and the backbone nitrogen of Asp368gp120. This hydrogen-bond occurs at the loop tip, an extra residue relative to CD4 is inserted in the strand, and the rest of the potential hydrogen bonds are of poor geometry or distance. Other similarities and differences with CD4 are found: of the two dominant CD4 residues (Phe43CD4 and Arg59CD4) involved in interaction with gp120, VRC01 mimics the arginine interaction, but not the phenylalanine one. Significant correlation is observed between gp120 residues involved in binding VRC01 and CD4. Superposition of the gp120 core in its VRC01-bound form with gp120s in other crystalline lattices and bound by other ligands indicates a CD4-bound structure (PDB ID 3JWD) (135) to be most closely related in structure, with a Cα-root-mean-square deviation of 1.03 Å. Such superposition of gp120s from CD4-bound and VRC01-bound conformations brings the N-terminal domain of CD4 and the heavy chain variable domain of VRC01 into close alignment, with 73% of the CD4 N terminal domain volume overlapping with VRC01 (131). This domain overlap is much higher than observed with the heavy chains of other CD4-binding site antibodies, such as b12, b13 or F105. In its complex with gp120, VRC01 rotates 43° relative to the CD4-defined orientation, and translates 6-Å away from the bridging sheet, to a clash-free orientation that mimics many of the interactions of CD4 with gp120, though with considerable variation. Analysis of electrostatics shows that the interactive surfaces of VRC01 and CD4 are both quite basic, though the residues types of contacting amino acids are distinct. Thus, while VRC01 mimics CD4 binding to some extent, considerable differences are observed.
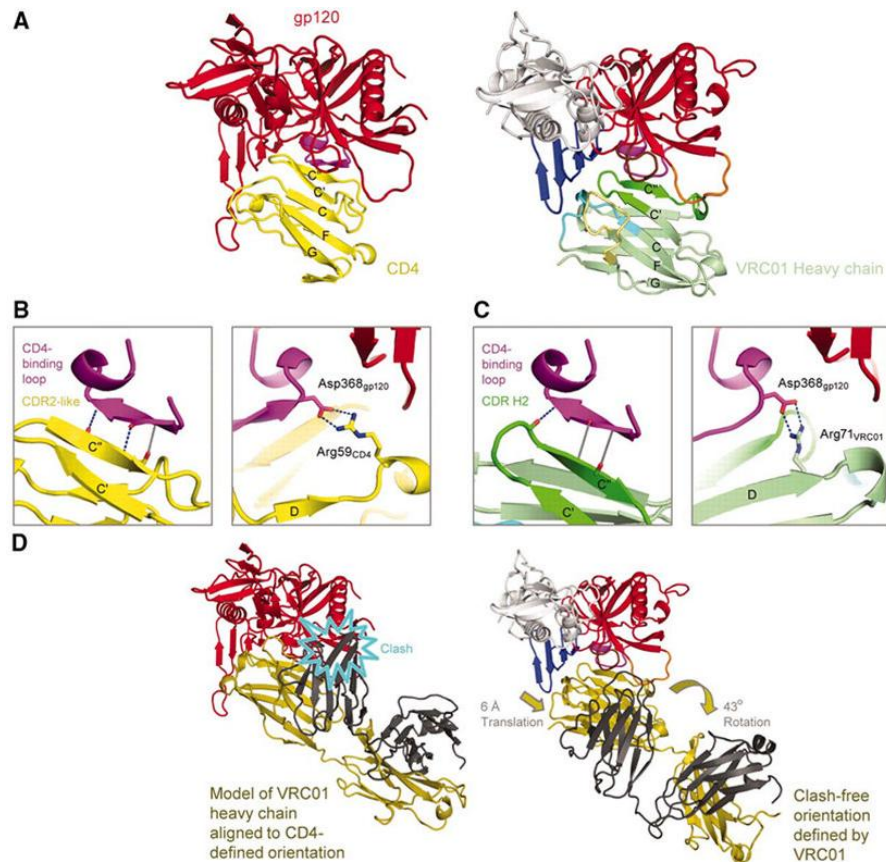
Figure 6: Structural mimicry of CD4 interaction by antibody VRC01

VRC01 shows how a double-headed antibody can mimic the interactions with HIV-1 gp120 of a single-headed member of the immunoglobulin superfamily, such as CD4. (A) Comparison of HIV-1 gp120 binding to CD4 (N-terminal domain) and VRC01 (heavy chain–variable domain). Polypeptide chains are depicted in ribbon representation for (right) the VRC01 complex and (left) the CD4 complex with the lowest gp120 root mean square deviation. The CD4 complex (3JWD) (16) is indicated in yellow for CD4 and red for gp120, except for the CDR-binding loop (purple). The VRC01 complex is colored as in Fig. 1. Immunoglobulin domains are composed of two b sheets, and the top sheet of both ligands is labeled with the standard immunoglobulin strand topology (strands G, F, C, C', C"). (B and C) Interface details for (B) CD4 and (C) VRC01. Close-ups are shown of critical interactions between the CD4-binding loop (purple) and the C" strand as well as between Asp368gp120 and either Arg59CD4 or Arg71VRC01. Hydrogen bonds with good geometry are depicted by blue dotted lines, and those with poor geometry are depicted in gray. Atoms from which hydrogen bonds extend are depicted in stick representation and colored blue for nitrogen and red for oxygen. In the left panel of (C), the b15-strand of gp120 is depicted to aid comparison with (B), although because of the poor hydrogen-bond geometry, it is only a loop. (D) Comparison of VRC01- and CD4-binding orientations. Polypeptides are shown in ribbon representation, with gp120 colored the same as in (A) and VRC01 depicted with heavy chain in dark yellow and light chain in dark gray. When the heavy chain of VRC01 is superimposed onto CD4 in the CD4-gp120 complex, the position assumed by the light chain evinces numerous clashes with gp120 (left). The VRC01-binding orientation (right) avoids clashes by adopting an orientation rotated by 43° and translated by 6 Å.

Natural resistance to VRC01

In addition to conformational masking and glycan shielding, HIV-1 resists neutralization by antigenic variation. Variation was observed in the V5 region in resistant isolates, and this variation – along with alterations in gp120 loop D - appeared to be the source of most

natural resistance to VRC01. Because substantial variation exists in V5, structural differences in this region might be expected to result in greater than 10% resistance. The lower observed frequency of resistance suggests that VRC01 employs a recognition mechanism that allows for binding despite V5 variation. Examination of VRC01 interaction with V5 shows that VRC01 recognition of V5 is considerably different from that of CD4, with Arg61VRC01 in the CDR H2 penetrating into the cavity formed by the V5 and β24-strands of gp120. Most importantly, the V5 loop fits into the gap between heavy and light chains; thus by contacting only the more conserved residues at the loop base, VRC01 can tolerate variation in the tip of the V5 loop.

To understand the structural basis for the exceptional breadth and potency of VRC01, we analyzed its interactive surface with gp120. VRC01 focuses its binding onto the conformationally invariant outer domain, which accounts for 87% of the contact-surface area of VRC01. The 13% of the contacts made with flexible inner domain and bridging sheet are non-contiguous and are not critical for binding. In contrast, CD4 makes 33% of its contacts with the bridging sheet, and many of these interactions are essential (134). The reduction in inner domain and bridging sheet interactions by VRC01 is accomplished primarily by a 6-Å translation relative to CD4 away from these regions; critical contacts such as made by Phe43 CD4 to the nexus of the bridging sheet-outer domain are not found in VRC01, while those to the outer domain (e.g. Arg59CD4) are mimicked by VRC01.

To determine the affinity of VRC01 for gp120 in CD4-bound and non-CD4-bound conformations, we used surface-plasmon resonance spectroscopy to measure the affinity of VRC01 and other gp120-reactive antibodies and ligands to two gp120s: a β4-deletion developed by Harrison and colleagues that is restrained from assuming the CD4-bound conformation (136) or a disulfide-stabilized gp120 core, largely fixed in the CD4-bound conformation in the absence of CD4 itself (134, 131). VRC01 showed high affinity to both CD4-bound and non CD4-bound conformations, a property shared by the broadly neutralizing b12 antibody (21-22). By contrast, antibodies F105 and 17b as well as soluble CD4 showed strong preference for only one, but not both, of the conformations. Thus, VRC01 interacts with the conformationally invariant outer domain in a manner that enables recognition of both CD4-bound and non-CD4-bound conformations of gp120.

Precise targeting by VRCO1

Prior analysis of effective and ineffective CD4-binding-site antibodies suggested that precise targeting to the initial site of CD4 attachment is required to block viral entry (137). The initial site of CD4 attachment is a subset of the CD4-binding site that involves the conformationally invariant outer domain (138).

Contacts by the VRC01 light chain (Tyr28VRC01 and Ser30VRC01) are made with the protein-proximal N-acetyl-glucosamine from the N-linked glycan at residue 276gp120 (131). Thus, instead of being occluded by glycan shielding, VRC01 makes use of a glycan for binding. Other potential glycan interactions may occur with different strains of HIV-1 gp120, as the VRC01 recognition surface on the gp120-outer domain extends further than that of the functionally constrained CD4-interactive surface, especially into the loop D and the often-glycosylated V5 region.
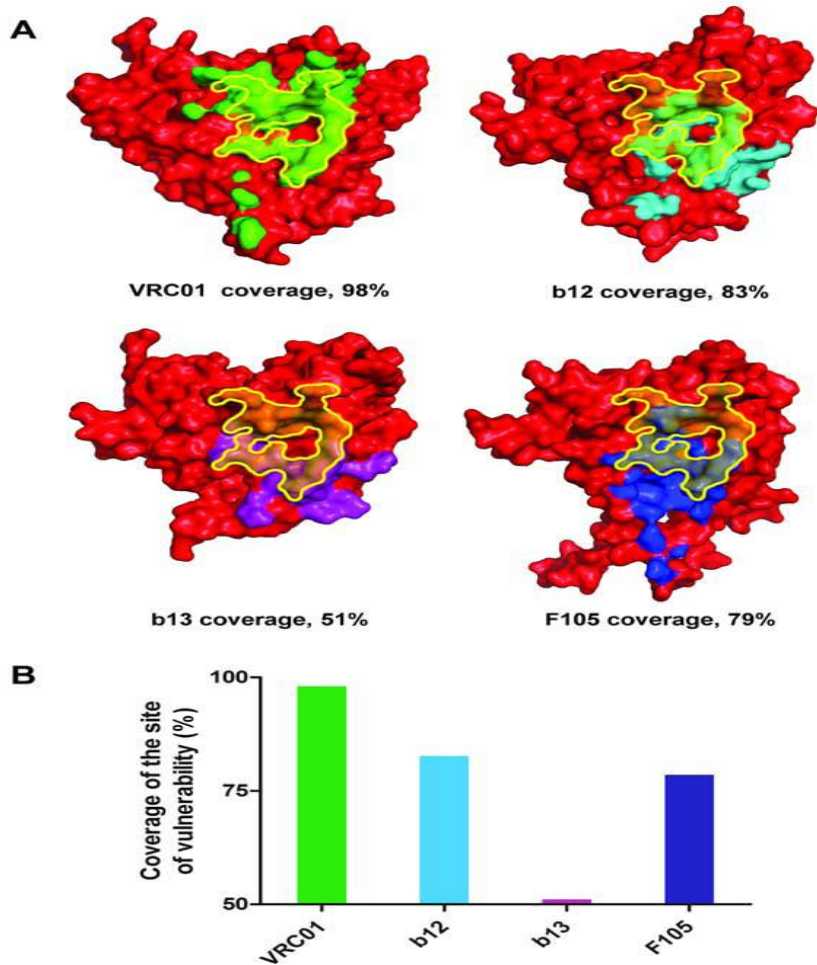
Figure 7: Comparison of coverage of the site of vulnerability
The site of vulnerability is the contact site for receptor CD4 on the outer domain of gp120. CD4-binding-site directed antibodies target this general area, however, most of them do not neutralize potently. When the site of vulnerability (semi-transparent yellow with solid yellow borderline) is superimposed over the antibody epitopes on gp120 surfaces (red), the degrees of overlapping differ. VRC01(green) hits the "bull's-eye" while b12 (cyan), b13 (purple) and F105 (blue) miss portions of the target with epitope straying away to other conformationally variable areas on gp120.
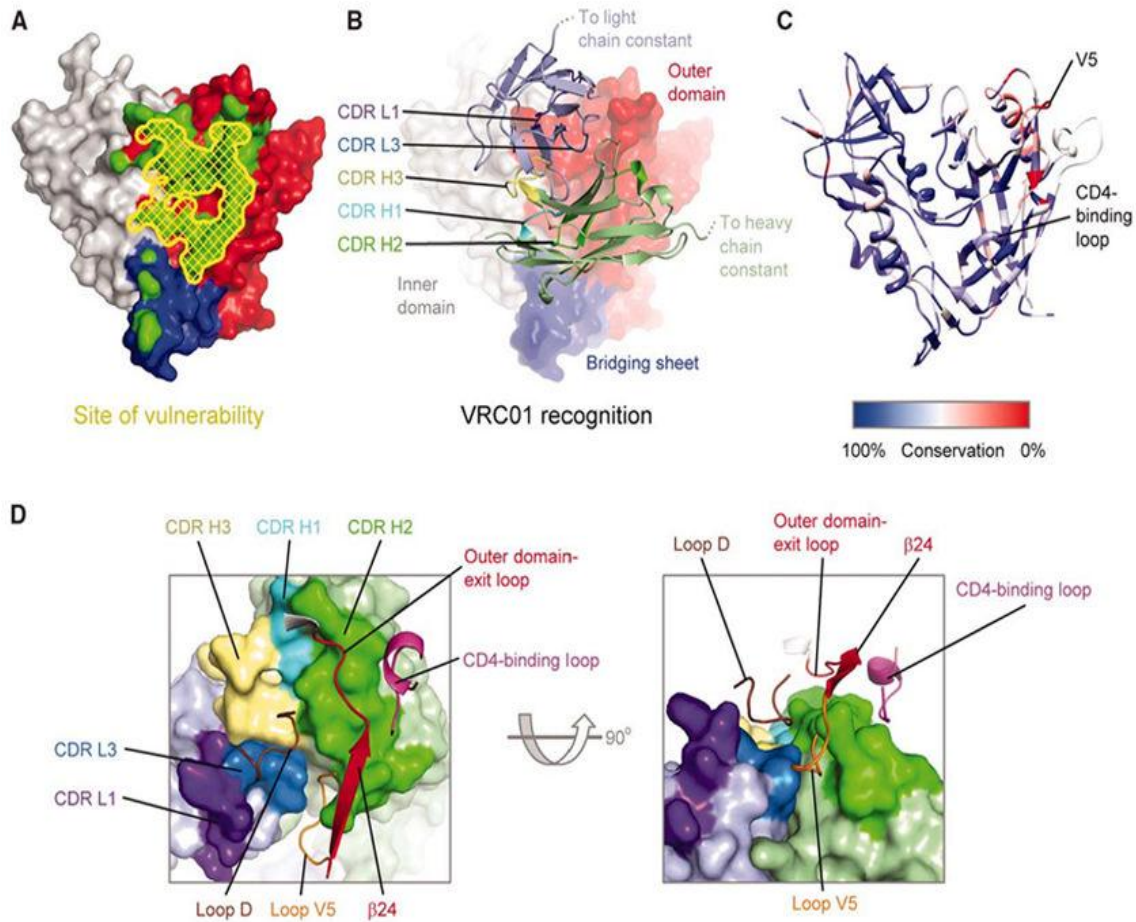
Figure 8: VRCO1 targeting
VRC01 precisely targets the CD4-defined site of vulnerability on HIV-1 gp120. Its binding surface, however, extends outside of the target site, and this allows for natural resistance to VRC01 neutralization. (A) VRC01 recognition and target site of vulnerability. The CD4 define site of vulnerability is the initial contact surface of the outer domain of gp120 for CD4 and comprises only 2/3 of the contact surface of gp120 for CD4 (22) the interactive footprint of VRC01 shown in green and the CD4-defined site of vulnerability outlined in yellow. (B) Antigenic variation. The polypeptide backbone of gp120 is colored according to sequence conservation, blue if conservation is high and red if conservation is low. (C) VRC01-resistant HIV-1 isolates. The 17 isolates of HIV-1 that resist VRC01 neutralization are displayed in ribbon and stick representation after threading onto the structure of gp120 in the VRC01-bound conformation. Side chains that clashed with VRC01 are highlighted in red. (D) Molecular surface of VRC01 and select interactive loops of gp120. Variation at the tip of the V5 loop is accommodated by a gap between heavy and light chains of VRC01.

REVIEW OF LITERATURE

AIDS vaccine: lessons learned, priorities and directions

The earliest phase I vaccine trials against HIV included recombinant envelope products, gp120 or gp160, formulated in various adjuvants. Live recombinant vectors such as vaccinia and canary pox and nucleic acid–based vaccines have also been tested, usually including gene inserts expressing the Env glycoprotein. Recombinant protein vaccines used alone, or as a boost to vaccine vectors, generally elicited high titers of anti-Env antibodies. Initial immunogenicity studies showed that vaccine-elicited antibodies could neutralize HIV-1 in vitro, but it was soon realized that the viral neutralization was limited to prototype laboratory-adapted HIV-1 strains and did not extend to primary HIV-1 isolates (139). These neutralization data generated considerable debate regarding the rationale for efficacy testing of protein-based vaccines.

Despite this uncertainty, two phase III gp120 vaccine trials were conducted, each with a bivalent formulation of two strains of gp120 formulated in Alum. The VAX004 and VAX003 studies were initiated in 1998 and 1999, respectively, and the results reported in 2003. These gp120 vaccines showed no significant impact on acquisition of HIV-1 infection and had no impact on plasma viremia or peripheral CD4þ T-cell counts (140; 141). The failure of these gp120 vaccines was generally viewed as evidence that a successful vaccine would need to induce more potent NAbs that can neutralize circulating strains of HIV-1.

The RV144 study included priming immunizations with an avipox vector (ALVAC) and boosting with the same bivalent gp120s used in the VAX003 study. The results of this study, released in 2009, showed that volunteers in the vaccine arm of the study acquired 31% fewer HIV-1 infections than those in the placebo arm (142). This modest efficacy, although not deemed adequate for licensure, was the first indication that a vaccine could protect against HIV-1 infection. The immunologic data from the RV144 vaccine trial are still being analyzed. Immune correlates analysis will include a set of in vitro antibody assays, including traditional virus neutralization, and assays of antibody binding, ADCC, and additional measures of Fc-mediated antibody effector functions.

The RV144 trial is not without its critics. The vaccine assessed in this trial did not stimulate broadly neutralizing antibodies (BnAbs) able to neutralize a broad range of transmitted or founder virus isolates8, generally regarded as the *sine qua non* of an HIV-1 vaccine, nor did it stimulate measurable responses by CD8+ cytolytic T cells4,5. Instead, the ALVAC-HIV/AIDSVAX B/E vaccine induced CD4+ T cell and antibody- dependent cell-mediated cytotoxicity responses and induced only neutralizing antibodies to the easy-to neutralize (tier 1) HIV-1 strains8. Haynes, Kim and colleagues9 coordinated a detailed attack on this problem by comparing a range of immune parameters in 41 vaccinated participants who became infected and 205 vaccinated subjects who did not become infected. Using robust assays, they found two strong correlates with infection risk. One was the plasma concentration of immunoglobulin G (IgG) antibody specific for the V1V2 loop region of envelope gp120, which was inversely correlated with infection risk. The other was high plasma concentrations of IgA antibody to HIV-1 Env, which were directly correlated with acquisition of infection. These findings have generated the following two hypotheses: that high concentrations of plasma antibodies specific for V1V2 are involved in protection against acquisition of HIV-1; and that high plasma concentrations of IgA to Env mitigate the effects of protective antibodies9. Many laboratories are now working to determine if those two correlates of risk in the RV144 trial are related mechanistically to the degree of protection noted in the trial or whether they are only surrogate markers for other factors. For example, if the types of V1V2-specific antibodies induced by the ALVAC-HIV/AIDSVAX B/E vaccine can be shown, after passive infusion into rhesus macaques, to protect against challenge with chimeric SHIV (simian immunodeficiency virus (SIV) with an HIV-1 envelope), then vaccines could be designed to induce V1V2- specific antibodies at concentrations higher than those seen in the RV144 trial. The ALVAC-HIV/AIDSVAX B/E vaccine simply decreased the chance of acquisition marginally; increasing the average number of exposures needed for infection and, in doing so, delayed infection. In this setting, a small vaccine-efficacy effect would show up in a low-risk cohort monitored very closely over time, but it might not show up in a high-risk sexual-transmission cohort or in high-risk intravenous drug users.   Post-hoc analysis of the vaccine-efficacy study of the VAX004 AIDSVAX B/B gp120 vaccine, which failed to protect Thai intravenous drug users against HIV infection, suggests that vaccine induced neutralizing antibody, CD4-blocking antibody

and antibody-dependent cellular viral inhibition were associated with lower rates of infection in vaccines (143).

Fortunately, there have also been exciting advances in the understanding of both BnAbs specific for HIV-1 and T cell immunity. HIV-1-specific monoclonal antibodies (MAbs) have been exceptionally useful reagents for characterizing neutralization-sensitive targets on the HIV-1 envelope glycoproteins and for revealing the structural dimensions of this complex heterotrimer.

One of the first MAbs to be extensively used for this purpose was IgG1b12, which was isolated in 1994 and targets the CD4 binding site (CD4bs). The CD4bs represents one of the most attractive targets for vaccine-induced immune responses because it is a highly conserved site present on all HIV-1 genetic subtypes as a result of the absolute requirement for engaging the CD4 molecule to gain entry into human cells.

Recently, VRC01, which is a considerably more potent anti-CD4bs MAb, has provided additional insights into the nature of this epitope (131). A cluster of new antibodies have helped to define a highly structural epitope present only on the trimeric envelope that includes conserved regions of both V2 and V3. The first identified MAb directed to this target, MAb 2090, was type-specific and neutralized only SF162 because this isolate carries an unusual lysine at position 160 in V2. PG9/16 isolated from a subtype A chronically infected individuals with exceptional plasma neutralization breadth binds the more common N160 variant and are hence more broadly cross-reactive (144).

The 2G12 MAb, which was isolated in 1996 (145), focused attention on glycans. The recent discovery of new MAbs that target the glycan at position 332 and compete with 2G12, but display more breadth and potency than 2G12, suggests that interest in this potential target is likely to increase (146). MAbs against the gp41 and the MPER have also recently been described and all define new epitopes in this region (147; 148).
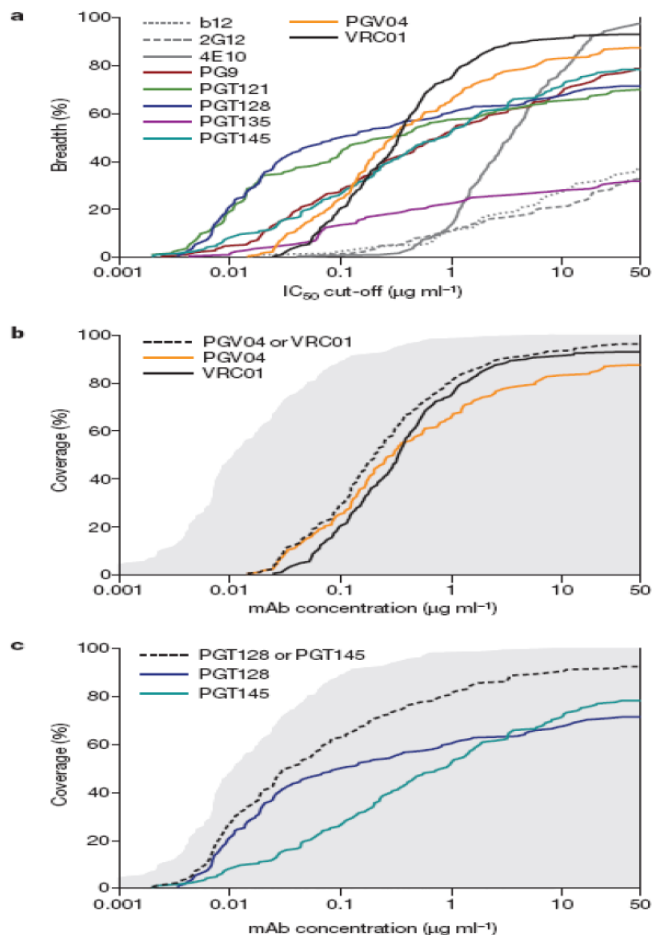
Figure 9: Certain antibody or antibody combinations are able to cover a broad range of HIV isolates at low, vaccine-achievable concentrations. a. Cumulative frequency distrubution of IC50 values of broadly neutralizing monoclonal antibodies tested against a 162-virus panel. The y-axis shows the cumulative frequency of IC50 values up to the concentration shown on the x-axis and can therefore also be interpreted as the breadth a specific IC50 cutoff. **b, c,** Percentage of viruses covered by single monoclonal antibodies (solid lines) or by at least one of the monoclonal antibodies in dual combinations of breadth (dashed black lines) dependent on individual concentrations. The grey area in both panels in the coverage of 26 monoclonal antibodies antibodies tested on the 162-virus panel (PGT121-123, PGT125-128, PGT130-131, PGT135-137, PGT141-145, PG-9, PG16, PGC14, VRC01, PGV04, b12, 2G12, 4E10, 2F5) and depicts the theoritical maximal achievable coverage known to date.

In summary, human efficacy trials indicate that vaccine-elicited protection against HIV-1 infection is achievable, but the specific antibody responses that may contribute to protection have not been elucidated. Despite these limitations in our knowledge of immune correlates, the modest 31% protection observed in the RV144 study suggests much room for improvement. One means of achieving improved vaccine efficacy may be through the elicitation of more potent and cross-reactive NAbs.

METHODOLOGY

DATABASES AND TOOLS

1. **Protein Databank (PDB)(** http://www.rcsb.org/pdb/home/home.do**)**

The RCSB Protein Data Bank (RCSB PDB) (http://www.pdb.org) is a member of the Worldwide Protein Data Bank (http://www.wwpdb.org). The wwPDB partners RCSB PDB (USA), PDBe (Europe, http://pdbe.org) (3), PDBj (Japan, http://www.pdbj.org) and BMRB (USA, http://www.bmrb.wisc.edu) act as data deposition, processing and distribution centers for PDB data. The PDB archive is the single world-wide repository of experimentally determined structures of proteins, nucleic acids, and complex biomolecular assemblies that is curated and annotated following standards set by the wwPDB. Each wwPDB partner offers unique views, query, analysis and visualization tools, and web services for the PDB archive on their respective web sites and databases. The RCSB PDB web site has undergone significant changes to improve usability, provide new query and analysis features, integrate additional external resources and enable user customization of the resource. The RCSB PDB web site caters to a wide variety of 'customers' from education (K-12, undergraduate, graduate), to academic and industrial researchers, to programmers and web developers. The redesigned web site supports the disparate requirements of this diverse user base. Major new or expanded features include new query and analysis tools, options to customize the web site, structural comparison of representative protein chains in the PDB, integration with literature from PubMed Central (http://www.ncbi.nlm.nih.gov/pmc) and binding affinity data from BindingDB (http://www.bindingdb.org). For web developers, new RESTful web services and web widgets have enabled the integration of RCSB PDB services and data into other web resources.

A detailed description how to use these widgets is also available. The RCSB PDB web site continues to take advantage of new scientific understanding and new technological developments. The powerful new Chemical Structure Search interface supports simple molecular weight, formula, substructure or similarity searches and complex SMARTS

queries. Faceted navigation significantly improves browsing query results with hierarchical navigation and the ability to refine a query iteratively based on new information gained during the search. Both the Chemical Structure Search and the new faceted search interface are tightly integrated with the 'Advanced Search' system to provide multiple paths for query refinement. The results of the queries can be tabulated, sorted, filtered and exported to enable large-scale data analysis. New sequence and structure analysis tools have been implemented. New Java versions of the FATCAT and CE algorithms for structural superposition have been made available for pair-wise alignments. In addition, all representative protein chains in the PDB have been structurally aligned. These pre-calculated alignments are updated weekly. Novel structures or new folds can be readily identified, as well as unexpected similarities between proteins of low-sequence identity. This information may be used to find evolutionary relationships or deduce previously unknown functions of proteins. PDB entries are now tightly integrated with the open-access literature that uses or cites the entries. The 'Literature View' of the RCSB PDB provides new ways to search and analyze the data. Data exchange and bi-directional links with BindingDB enable correlation of structure with binding affinity data.

The 'What's New' page (http://www.pdb.org/pdb/static.do?p=general_information/whats_new.jsp) lists detailed summaries of the latest and past improvements. The addition of these new features and improvements represents a new generation of the RCSB PDB web site. It allows more complex analysis to be performed and provides systematic comparisons across all of the PDB with the goal to further scientific discovery and education. [144]

2. Rosetta Multigraft Match (http://www.rosettacommons.org/)

The matching algorithm was developed to select scaffolds suitable for transplantation of structural motifs, based on structural compatibility with a given motif and on their predicted ability to interact with a desired binding partner. The starting input for Multigraft Match is the set of atomic coordinates of the motif in complex with the desired binding partner (or if no partner is available, coordinates of the motif alone can be used). The matching of a multi-segment motif is summarized in the following stages: I) Primary

Loop Matching; II) Secondary Loop Matching; III) Binding Partner Clash check. The algorithm allows the identification of protein scaffolds for single or multi-segment structural motifs.

Primary loop matching: Each backbone segment of the complex motif is considered as the Primary loop. The Primary loop is excised from the remaining segments of the input motif and then aligned on every residue of all the scaffolds contained in a culled PDB to identify potential sites where it can be transplanted.

Different types of alignment systems like superposition alignment, N2C alignment, C2N alignment and E alignment systems are used. From the four alignment systems, only superposition evaluates the local similarity of the existing backbone comprehensively. The C-terminus and N-terminus alignments are conceptually identical, differing only by which terminus of the Primary Loop is aligned onto the scaffold.

The C2N and N2C alignments are further divided into different subsets: N centered (C2N_N), C centered (N2C_C), Cα centered depending on which atom is used for aligning primary loop to the residues on the scaffold. The End-point alignment superimposes the three backbone atoms (N, Cα, C) of both Primary loop termini onto the corresponding atoms of residue pairs on the query scaffold.

Two different types of filters are used for selecting the potential matches: geometric and steric clash

The geometric filter for the superposition alignment evaluates the backbone root mean square deviation (rmsd) of the Primary Loop relative to a segment in the scaffold. In the other three alignment types (C2N, N2C, E) the geometric filter evaluates the rmsd between the termini of the Primary loop and proximal scaffold residues to yield a chain-break score. The steric-clash filter ensures that the particular alignment of the Primary loop on the scaffold does not clash with other backbone regions of the scaffold ("intra-clash"). Matches will be carried forward if both filters are satisfied within user defined thresholds. Typically, the matching procedure is carried out with successively looser thresholds for geometrical and steric filters, to ensure that all promising matches are identified efficiently.

Secondary loop matching: After satisfying the Primary loop matching filters, the remaining segments of the input complex motif are tested as Secondary loop matches. Secondary loop matches are tested as End-point matches. Each candidate secondary loop segment is recovered in the context of the scaffold while retaining the original orientation relative to the Primary loop. The Secondary loop alignment has to satisfy both the geometric and steric clash filters. In practice the user-defined filter thresholds are typically less stringent for secondary matches. If both filter criteria are satisfied, the match proceeds to the next stage.

Clash check: In the final stage of the matching protocol the orientation of the input complex is recapitulated between the binding partner and a "hybrid" scaffold structure composed of the scaffold with the input motif in the match-derived orientation. The theoretical binding energy of the complex is computed and matches with considerable steric clashes across the interface ("inter-clash") are discarded.

3. **Rosetta Multigraft Design(**http://www.rosettacommons.org/**)**

Multigraft Design was implemented to handle the Multigraft Match output automatically, to keep track of the mapping between the matched motif residues and the scaffold. It is composed of several stages I) Match reconstitution; II) Build connecting segments by fragment insertion; III) Sequence design; IV) Filtering; V) Human-guided design.

In the first stage of match reconstitution, a hybrid structure of scaffold with the input motif in the matched rigid body orientation is assembled. These hybrid structures will have one or more discontinuities where the backbone of the scaffold and the epitope are disconnected. To close the gaps, linkers are added as per the user defined parameters. This is done by a backbone building algorithm that performed fragment insertion and Cyclic Coordinates Descent Steps. The length of the linkers and the secondary structure of the linkers to be built are user specified. The backbone conformations satisfying the chain break thresholds are processed further. In the sequence design stage, the newly built linkers are further optimized by sequence modifications followed by multiple rounds of

side chain and backbone minimization. At the interface between the antibody and regions of the scaffold outside the transplanted epitope, the scaffold residues are mutated to eliminate any potential steric clashes or nonspecific interactions – the default allowed residues at these positions are A, G, S, and T. All the steps in this stage are carried out using a full-atom energy function for accurate energetic assessment. The filtering stage ranks and selects the designs with lower Rosetta energy; it removes the designs with structural flaws. Lastly, the human guided design stage inspects and rectifies the aspects that were not considered in the automated stage like to remove unpaired cysteines, functional sites, oligomerization interfaces and also to mutate solvent exposed hydrophobic residues to enhance the protein solubility.

4. **Pymol(**http://www.pymol.org/**)**

PyMOL is an open-source, user-sponsored, molecular visualization system created by Warren Lyford DeLano and commercialized by DeLano Scientific LLC, which is a private software company dedicated to creating useful tools that become universally accessible to scientific and educational communities. It can produce high quality 3D images of small molecules and biological macromolecules, such as proteins. According to the author, almost a quarter of all published images of 3D protein structures in the scientific literature were made using PyMOL.

PyMOL is one of a few open source visualization tools available for use in structural biology. The **Py** portion of the software's name refers to the fact that it extends, and is extensible by the Python programming language.

It can be run interactively (using menu-driven commands and options) or *via* input scripts/macros, but most efficiently by a combination of both approaches. It can run on Linux, Windows and OSX.

5. **Computational Alanine Scanning Server(**http://robetta.bakerlab.org/alascansubmit.jsp**)**

Experimental —alanine-scanning mutagenesis is a powerful method for analyzing important interactions in protein-protein interfaces. Alanine scanning measures the effect

of the deletion of an amino acid side chain beyond the Cβ carbon atom on the affinity of a protein-protein complex. Individual substitutions of many amino acids with alanine yield a map of which interactions are critical in an interface and which ones are not.

Clackson and Wells called these energetically important residues —hot spots in their pioneering work on the binding of human growth hormone to its receptor, where only a small fraction of interface residues account

for the majority of the binding energy. Subsequent studies suggested that the presence of binding energy hot spots comprising only a fraction of the complete interface area is a general property of most protein-protein complexes.

Computational alanine scanning uses a simple free energy function to calculate the effects of alanine mutationson the binding free energy of a protein-protein complex. The function consists of a linear combination of a Lennard-Jones potential to describe atomic packing interactions, an implicit solvation model, an orientation-dependent hydrogen-bonding potential derived from high-resolution protein structures, statistical terms approximating the backbone-dependent amino acid-type and rotamer probabilities, and an estimate of unfolded reference state energies

If computational alanine scanning is applied to a complex structure without existing experimental data, the algorithm automatically identifies all interface residues in a protein-protein interface. An interface residue is defined as (i) a residue that has at least one atom within a sphere with a 4 Å radius of an atom belonging to the other partner in the protein complex, or (ii) a residue that becomes significantly buried upon complex formation, as measured by an increase in the number of Cβ atoms within a sphere with a radius of 8 Å around the Cβ atom of the residue of interest. The program then replaces each of the interface residues individually with alanine residues, and computes the effect of this mutation on the binding free energy of the complex.

Hot-spot residues can be defined operationally as those for which alanine mutations have destabilizing effects on ΔΔGbind of more than 1 kcal/mol. For a comparison with experimental data, a correctly identified hot spot means a residue with a predicted and

observed ΔΔGbind value of greater than or equal to 1 kcal/mol, and a correctly identified neutral residue has both predicted and observed ΔΔGbind values of less than 1 kcal/mol.

Several assumptions are made in implementation of computational alanine scanning that affect the applicability of the method and the interpretation of the results.

First, the terms in the energy function are pairwise additive. This simplification is necessary for a fast computational evaluation of residue-residue energies used in protein design methods. Because of this approximation, coupling effects between different mutations cannot be taken into account, and multiple mutations are always assumed to be additive. Thus, experimental data on non additivity effects observed in double-mutant cycles cannot be reproduced by these methods. For the same reason, indirect effects on the binding energy exerted by residues not making direct interactions in the interface are generally not captured.

Second, the isolated partners in the protein-protein complex have the same bound and unbound conformations during the calculations; it is assumed that there are no conformational changes upon binding. Consequently, intramolecular interactions are equivalent in the isolated complex partners and in the complex, and their contribution cancels out in the calculation of the binding energy. The only exceptions are electrostatic interactions that are considered to be dependent on their environment. For example, an intramolecular hydrogen bond can become more buried in the complex than in the monomer, and is therefore predicted to be stronger, according to the environment-dependent energy function. Thus, if one of the side chains participating in the hydrogen bond is mutated to alanine, there will be a net destabilizing effect on the binding free energy even if the hydrogen bonding interaction is not intermolecular. For the same reason, residues not directly participating in contacts across the interface can affect the binding free energy by changing the environment of hydrogen-bonding interactions.

Third, all structures are viewed as static; possible changes in side-chain and backbone mobility of the interacting proteins are not taken into account. Ignoring these effects can lead to errors, especially where disordered regions become ordered upon binding [some

protein parts may also become more flexible upon binding]. The entropy gains and losses associated with these effects can significantly affect the binding free energy.

Fourth, cofactors, metal ions, hydrogen-bonding water molecules bridging side-chains in the protein interface, or other nonpeptide ligands or binding partners (such as nucleic acids) are not taken into account. This approximation will lead to an underestimation of the energetic effects of mutating residues contacting these nonpeptide ligands.

Fifth, symmetry in a protein-protein complex is not taken into account. During computational alanine scanning, only one residue at a time is considered. Therefore, in a symmetrical interface with one mutated position, the corresponding residue on the other partner in a dimer is modeled as wild type. However, because of the assumption of additivity, the single mutations of corresponding residues can be added to a first approximation.

6. Protein Interaction Calculator Server (http://pic.mbu.iisc.ernet.in/)

PIC (Protein Interactions Calculator) is a web based service to aid recognition and analyses of various kinds of interactions in tertiary structures of proteins and structures of protein–protein complexes. Solvent accessibility calculations have also been integrated in PIC to aid recognition of interacting motifs that are exposed or buried. Further, the residue depth calculations are also made possible in PIC so that interactions deep inside the protein structure or near the surface can be recognized. Advantage of using residue depth parameter is that it can distinguish residues with no solvent accessible surface area in terms of how deep they are from the protein surface.

PIC server accepts atomic coordinate set of a protein structure in the standard Protein Data Bank (PDB) format. The user is prompted with selecting one or more of the following interaction types: Interaction between apolar residues, disulphide bridges, hydrogen bond between main chain atoms, hydrogen bond between main chain and sidechain atoms, hydrogen bond between two sidechain atoms, interaction between oppositely charged amino acids (ionic interactions), aromatic–aromatic interactions, aromatic–sulphur

interactions and cation–p interactions. The input coordinate set is accepted, under each section of the page, for recognition of interactions within a polypeptide chain. If an ensemble of NMRderived structures is input then the first model in the file is taken as a representative and is used by the PIC server. The output corresponds to the list of residues involved in interaction type of interest. An option is provided, using RasMol interface and Jmol interface, for enabling visualization of structure in the graphics with interactions highlighted. It is possible to get the results by e-mail. It is also possible to download the output files of the original programs. Solvent accessibility calculations could be used to identify different kinds of interactions between buried or between solvent exposed residues. Solvent accessibility calculations are performed using NACCESS program (Hubbard, S.J. and Thornton, J.M., 1993, NACCESS Computer Program, Department of Biochemistry and Molecular Biology, University College London.). The exposed and buried residues are identified by >7% and <=7% residue accessibility, respectively. Under this facility list of all the interaction types are displayed prompting the user to select list of interaction types of interest.

Depth of an atom in a protein is defined as the distance from the nearest atom in the surface of the protein structure. Mean depths of atoms of a residue defines the residue depth. Analogous to the panel on solvent accessibility, panel on residue depth enables the users to identify specific types of interactions near the protein surface or deep inside the core of the structure. Those residues with depths <=5A° are considered as close to the protein surface and others as deep inside. Using this part of the PIC server it is possible to identify interactions between, say, aromatic residues near the protein structural surface. As calculation of residue depths takes a few minutes for most protein structures, results involving depth calculation are sent by e-mail to the user if a valid e-mail address is provided.

Criteria used by PIC webserver for recognizing various types of interactions

1. HYDROPHOBIC INTERACTIONS:

The following residues are considered to participate in interactions if they fall within 5Å range.
ALA, VAL, LEU, ILE, MET, PHE, TRP, PRO, TYR.

Reference: Kyte and Doolittle.

2. DISULPHIDE BRIDGES:
Pairs of cysteines (sulphur atoms) within 2.2 Å are considered as disulphide bridges.

3. HYDROGEN BOND :
Criteria for hydrogen bond definition :
donor acceptor distance cutoff (oxygen and nitrogen) = 3.50
donoracceptor distance cutoff (sulphur) = 4.00
Reference: J. Overington, et al. *Proc. Roy. Soc. Biol Sci.* 1990, pg.132145

4. IONIC INTERACTIONS:
Ionic residue (ARG,LYS,HIS,ASP,GLU) pairs falling with 6Å (default) contribute to ionic interactions.

5. AROMATIC-AROMATIC INTERACTIONS:
Pairs of phenyl ring centroids that are seperated by a preferential distance of between 4.5 to 7 Å accounts for aromatic interactions.
Reference: S.K.Burley, G.A.Petsko, *Science*, 1985, Vol 299, pg.2328.

6. AROMATIC-SULPHUR INTERACTIONS:
Interactions between the sulphur atoms of cysteine and methionine and the aromatic rings of phenylalanine, tyrosine and tryptophan within 5.3 Å account for aromatic sulphur interactions.
Reference: K.S.C Reid, P.F.Lindley and J.M. Thornton, *FEBS letter* 1985, Vol.190, pg.209213.

7. CATION PI INTERACTIONS:
When a cationic side chain (Lys or Arg) is near an aromatic side chain (Phe, Tyr, or Trp) within 6 Å separation they account for cation pi interactions.
Reference: R.Satyapriya and Saraswathi Vishveshwara, *Nucleic Acid Research* , 2004, Vol 32, pg.41094118.

8. SOLVENT ACCESSIBILITY:
To calculate the total accessible surface area of a protein NACCESS program is used.
Residues that have a total relative accessibility value greater than 7 are considered as exposed.
Residues that has a total relative accessibility value less than or equal to 7 are considered as buried.
Reference: S.J. Hubbard, & J.M. Thornton, 'NACCESS'. 1993. (VERSION 2.1.1).

9. DEPTH CALCULATION:
Atoms near surface are recognised with residue depth value less than or equal to 5.
Reference: Suvobrata Chakravarty and Raghavan Varadarajan, *Structure*, 1999, Vol 7, Pg.723–732.

7. CEP (http://pic.mbu.iisc.ernet.in/)

CEP server provides a web interface to the conformational epitope prediction algorithm developed in-house. The algorithm, apart from predicting conformational epitopes, also predicts antigenic determinants and sequential epitopes. The epitopes are predicted using 3D structure data of protein antigens, which can be visualized graphically. The algorithm employs structure-based Bioinformatics approach and solvent accessibility of amino acids in an explicit manner. Accuracy of the algorithm was found to be 75% when evaluated using X-ray crystal structures of Ag–Ab complexes available in the PDB.

The CEP is implemented on Apache server with Linux 9.2 as an operating system. The web interface is designed using CGI Perl and JAVA scripts. The visualization package, Jmol, which is an open source software suite (http://jmol.sourceforge.net/), has been plugged in to facilitate visualization of the predicted conformational and sequential epitopes. Results are displayed in html format.

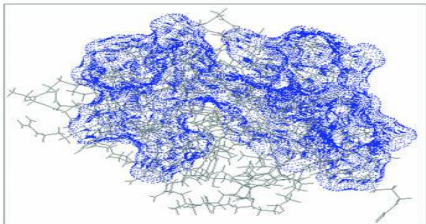CEP server requires the 3D coordinate data in PDB format.

Output is generated in html format. Predicted AD and CE are listed as separate tables.

AD number is followed by the chain ID and amino acid sequence of predicted AD along with the start and end positions. The residues that satisfy the percentage accessibility criterion are shown in uppercase, whereas those that do not satisfy this criterion are shown in lowercase. They are part of AD as they fulfil the criteria for extension (see Algorithm). The reference AD for every CE is shown in green and individual residues that are part of CE and SE are also listed. The predicted AD, CE and SE can be visualized by clicking the respective radio buttons. The experimentally characterized BS for the evaluation dataset can be visualized by clicking the BS radio button.

Example output:

The CEP algorithm has been evaluated using a curated dataset consisting of 21 co-crystal complexes available in PDB. In the process of evaluation, it was observed that the algorithm predicts relatively larger binding sites for a few antigens, which may appear as false-positive predictions. An explanation for this can be drawn from the analyses of structures of antigens with multiple antibodies, such as lysozyme and neuraminidase. Complexes of lysozyme with various Abs have shown that the Abs of lysozyme have overlapping binding sites.

Furthermore, it has also been shown that the same residue of an antigen may be a part of different epitopes and interact in a unique manner with respective paratopes. Thus, the residues that may appear 'additional' in the predicted CE need not be referred to as false positives, since the algorithm predicts all possible binding sites of the given antigens. Hence, the predicted CE/Ag-binding site is the sum of the binding sites of the individual antibodies.

Visualization and mapping of the predicted ADs and CEs on a given 3D structure enhances utility of the server. It must be mentioned that the usability of this server is limited by the availability of the 3D structure data of protein antigens.

8. Discotope (http://www.cbs.dtu.dk/services/DiscoTope/)

DiscoTope uses a combination of amino acid statistics, spatial information and surface exposure. It is trained on a compiled data set of discontinuous epitopes from 76 X-ray

structures of antibody/antigen protein complexes. Prediction of B-cell epitopes may help to identify epitopes in proteins that have been analyzed using experimental techniques based on antibody affinity binding, e.g., Western blotting, immunohistochemistry, radioimmunoassay (RIA), and enzyme-linked immunosorbent assay (ELISA). Combination methods using a number of propensity scales have been used for B-cell epitopes for more than 15 years (Pellequer et al. 1991); however, DiscoTope is the first reported method combining a propensity scale with three-dimensional structural information, such as spatial proximity.

DiscoTope detects 15.5% of residues located in discontinuous epitopes with a specificity of 95%. At this level of specificity, the conventional Parker hydrophilicity scale for predicting linear B-cell epitopes identifies only 11.0% of residues located in discontinuous epitopes.

In order to use the server for prediction of Bcell epitopes, any of the following three ways of submission can be performed:

1. Write a PDB entry name and chain id(s) into the corresponding windows.
2. Specify a file from your local disk containing a list of existing PDB entries with specified chain IDs, one per line, in the format 'entryname_chain' e.g. **1zz6_B**
3. Specify a file from a local disk, in PDB format, and chain id(s).

Discotope also allows the selection of threshold score for epitope identification.

Different thresholds for the DiscoTope score can be translated into sensitivity/specificity values. In a benchmark containing more than 75 antigen/antibody complexes, the following relations were found:

Score	Sensitivity	Specificity
> -3.1	0.16	0.95
> -4.7	0.24	0.90
> -6.0	0.32	0.85
> -6.9	0.40	0.80
> -7.7	0.47	0.75

A specificity of 0.75 would mean that 25% of the nonepitope residues were predicted as part of epitopes.
A sensitivity of 0.47 would mean that 47% of the epitope residues were predicted as part of epitopes.

The output consists of eight columns namely chain id, residue number, amino acid name, contact number, propensity score, discotope score and <=B. Identified B cell epitope

Example output:

DiscoTope predictions for '1zz6'.
Looking only at Chain: A
Contact Distance = 10.000 Angstroms
Threshold = -7.700

```
A        6        THR      16       -5.334    -13.334
A        7        ALA      25       -9.276    -21.776
A        8        SER      22       -7.223    -18.223
A        9        THR      19       -6.091    -15.591
A        10       GLY      16       -6.217    -14.217
A        11       PHE      27       -8.426    -21.926
A        12       ALA      25       -7.100    -19.600
A        13       GLU      18       -5.335    -14.335
A        14       LEU      18       -4.040    -13.040
A        15       LEU      24       -5.943    -17.943
A        16       LYS      20       -2.846    -12.846
A        17       ASP      14       -0.583    -7.583     <=B
A        18       ARG      20       -1.241    -11.241
A        19       ARG      21       -2.802    -13.302
A        20       GLU      14       0.918     -6.082     <=B
```

9. SEPPA(http://lifecenter.sgst.cn/seppa/)

In recent years, a lot of efforts have been made in conformational epitope prediction as antigen proteins usually bind antibodies with an assembly of sequentially discontinuous and structurally compact surface residues. In the method of SEPPA, a concept of 'unit patch of residue triangle' was introduced to better describe the local spatial context in protein surface. Besides that, SEPPA incorporated clustering coefficient to describe the spatial compactness of surface residues.

SEPPA was rigorously trained by 82 antigen–antibody protein complexes, which contained 84 unique epitopes. One hundred and nineteen independent spatial epitopes of protein antigens were collected as testing dataset. Only those with resolution better than 3.0 Å and protein antigen length with more than 25 residues were used. Redundant epitopes were removed by 60% similarity. The testing data were collected from the training dataset of DiscoTope, databases of IEDB and Epitome. A testing dataset of 119 antigens was set up by deducting SEPPA's training data from above testing data.

Validated by independent testing datasets, SEPPA gave an average AUC value over 0.742 and produced a successful pick-up rate of 96.64%. Comparing with peers, SEPPA shows significant improvement over other popular methods like CEP, DiscoTope and BEpro. In addition, the threshold scores for certain accuracy, sensitivity and specificity are provided online to give the confidence level of the spatial epitope identification. The web server can be accessed at http://lifecenter.sgst.cn/seppa/index.php. Batch query is supported.

Definition for Unit patch of a residue triangle: Solvent accessible surface areas (SASA) are determined (Naccess V2.1.1.) for each residue in antigen proteins. Surface residues are those with more than 1 Å^2 SASA, while those with SASA loss in binding of more than 1 Å^2 are classified as epitope residues. The unit patch of residue triangle is defined among any three surface residues if the distance for every two of them is within 4 Å atom distance. Based on the training data, unit patches containing more than two epitope residues are termed as epitope unit patches; otherwise are classified as non-epitope unit patches.

Definition of residue neighbor and clustering coefficient: Clustering coefficient is introduced to describe the compactness of the neighboring residues around one residue. It reflects the probability that the neighbors of residue r are also neighbors with each other. For one residue r, all residues within 15 Å of r are defined as residue neighbors of r. k_r is the total number of residues neighbors for r. Theoretically, the number of all possible links among k_r residue neighbors is $k_r(k_r-1)/2$. However, as only those links within certain distance can be called residue neighbors of each other, the observed actual number of residue neighbor pairs among k_r is recorded as e_r. The clustering coefficient (cc_r) is given as :

$$cc_r = \frac{e_r}{[k_r(k_r-1)/2]}.$$

SEPPA requires a 3D protein structure in PDB format as input. Users can submit the query with a released PDB ID or upload a structural file in PDB format. It is recommended to specify the chain(s) ID if not all peptides are antigen proteins to be queried in the structure file. Otherwise, each chain will be assumed as an antigen protein and calculated for antigenicity scores.

The results of prediction are displayed in html format. The sequence of submitted protein antigen is displayed in single letter code in result window. The core residues are shown in lowercase and surface residues in uppercase. The residues predicted as epitope are highlighted with yellow color background. The scores of prediction are recorded in another file, which lists the antigenicity scores for individual residue and this file is downloadable. A link to visualize the prediction result is also provided in the result page. The visualization of result is displayed with Jmol (an open-source Java viewer for chemical structures in 3D). Tints from blue to red represent a rising propensity for a residue to be in the epitope.

10. PEPITO(http://pepito.proteomics.ics.uci.edu/info.html)

PEPITO, attempts to overcome some of the limitations of previous predictors by incorporating an amino acid propensity scale along with side chain orientation and solvent accessibility information using half sphere exposure values. To increase robustness, PEPITO uses propensity scales and half sphere exposure values at multiple distance thresholds from the target residue.

For each residue r in the target protein chain, an epitope score E(r) is calculated. Large values of E(r) indicate a higher likelihood that the residue r is an epitope residue. The score E(r) is calculated using a linear combination of terms. Non-linear methods such as SVMs, ANNs and Gaussian Mixture Models were also explored, but they did not achieve higher performance levels. The score is given by:

$$E(r) = \sum_{k \in \{8, 10, \ldots, 16\text{Å}\}} \alpha \cdot PS(r, k) + \beta \cdot \text{HSEup}(r, k) + \gamma \cdot \text{HSEdown}(r, k)$$

The first term PS(r, k) is the sum of the propensity scale scores, averaged over a linear window of nine residues, for all residues within k A ° of residue r. The second half-sphere exposure term HSEup(r, k) is the number of $C\alpha$ atoms in the up half sphere within k A ° of residue r, and similarly for the third term using the down half sphere within k A ° of residue r. Intuitively, the HSEup term encodes information on the relative orientation of the side chain—toward the center of the protein or toward the surface—and the side chain accessibility.

Currently, PEPITO uses the propensity scale. The coefficients $(\alpha=1, \beta=-1/2, \gamma=-1/4)$ are derived from those previously used by Andersen et al. (2006) and the correlations between half sphere exposures and contact number.

The server version of PEPITO calculates the epitope score using all residues—only the antigen chain should be used, not the antibody–antigen complex. PEPITO returns a simplified PDB file with the epitope score, expressed as a Z-score, in the B-factor field of each atom. A Z-score threshold of 1.3 will produce a sensitivity>0.3 and specificity>0.9. On the Discotope dataset, PEPITO achieves an ROC AUC of 75.38 while on the Epitome dataset, PEPITO achieves an ROC AUC of 68.31.

WORKFLOW

A flowchart showing the workflow:

- **START** (green)
- **DEFINE THE EPITOPES** (LIT. REFERENCES, EPITOPE PREDICTION TOOLS AND COMP ALA SCAN) (orange)
- **Culled PDB** (orange cylinder)
- **SELECT THE SCAFFOLDS** (ROSETTA MATCH) (yellow)
- **GRAFT THE DISCONTINUOUS EPITOPE ON THE SCAFFOLD** (ROSETTA MULTIGRAFT DESIGN) (peach)
- **ENERGY MINIMIZATION OF THE SELECTED DESIGNS** (ROSETTA FASTRELAX) (green)
- **ANALYSE THE STRUCTURES** (RMSD, PIC AND PISA WEBSERVER) (blue)
- **ARE THE STRUCTURES GOOD?** (pink diamond) — NO / YES
- **STOP** (red)

1. DEFINING THE EPITOPE

The discontinuous epitope was selected as per the following criteria:
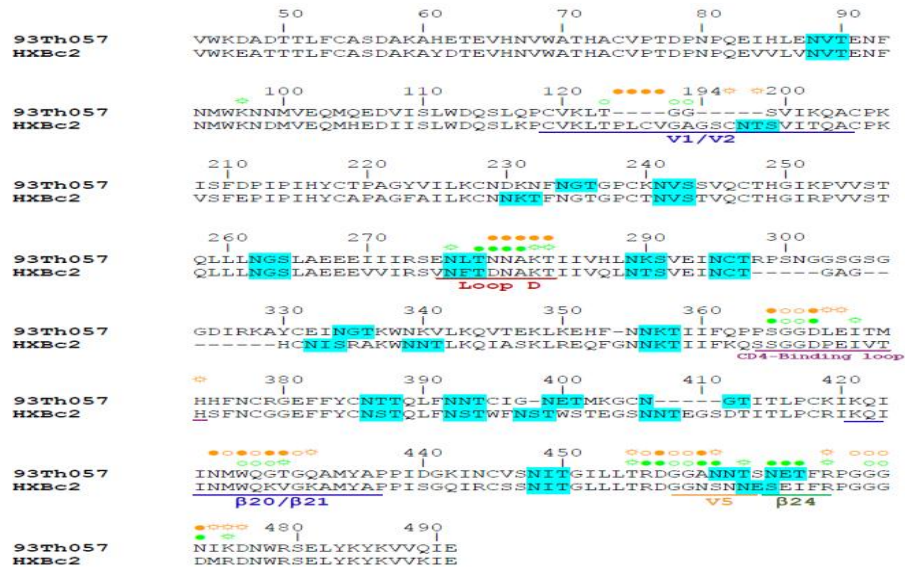
a. Literature references

Figure 10 : gp120 contacts as defined with the program PISA for the CD4 and VRC01 complexes are indicated in orange and green, respectively, with Open circles (○) denoting gp120 main-chain-only contacts, Open circles with rays (☼) denoting gp120 side-chain-only contacts, and Filled circles (●) denoting both main-chain and side-chain contacts.

The major structure elements of gp120 involved in ligand binding are underlined.

b. Results of computational alanine scanning: computational alanine scanning was performed for the complex G.pdb (gp120-VRCO1 complex). The scan identified the hotspot residues in the complex. The scan was performed to identify the discontinuous motif that could be a possible target for designing immunogens.

Hotspots: hotspot residues can be defined operationally as those for which alanine mutations have destabilizing effects on ΔΔGcomplex of more than 1kcal/mol.

The results suggested that residues within the loopD and CD4bs were important for gp120-VRCO1 interaction. Most of the hotspots lied within these regions.

c. **Results from various epitope prediction servers**
The following discontinuous epitope prediction servers were used to further support the selection of our epitopes:

Discotope(http://www.cbs.dtu.dk/services/DiscoTope/)
SEEPA(http://lifecenter.sgst.cn/seppa/index.php)
PEPITO/BEPRO(http://pepito.proteomics.ics.uci.edu/info.html)
CEP(http://bioinfo.ernet.in/cep.htm)

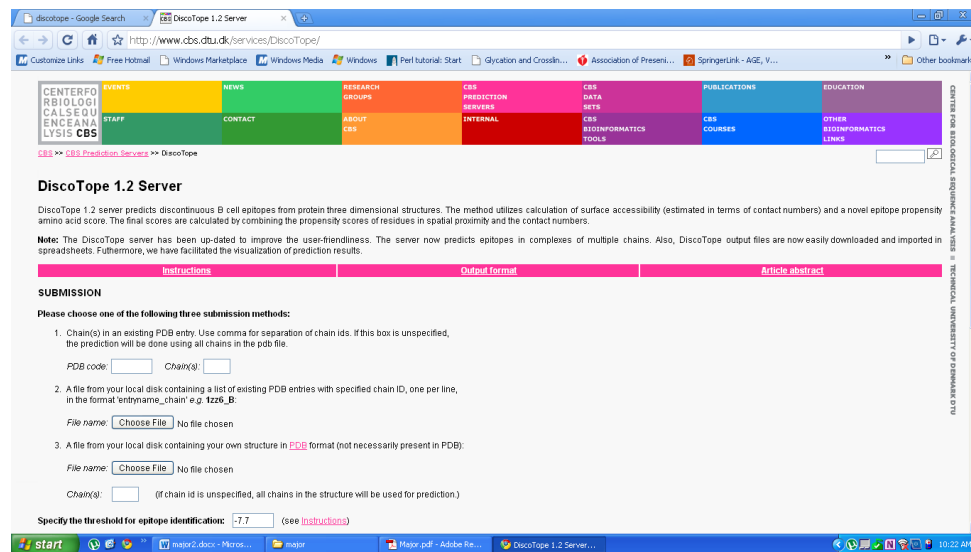The results of the above mentioned prediction server was combined to pick out only few highly probable discontinuous epitope regions.

Predictions using Discotope

1). Homepage of Discotope webserver was opened at
http://www.cbs.dtu.dk/services/DiscoTope/

2). After going through the instructions the PDB file was uploaded. All other parameters were taken at their default values and submitted.
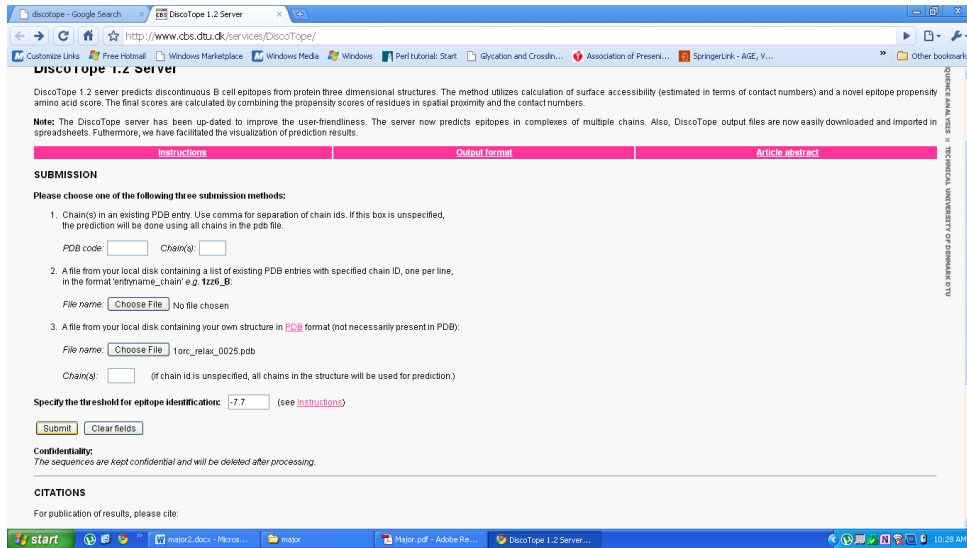
3). After pressing the submit, user is directed to a page asking the user to enter the email address to which the results have to be mailed (appendix2).

Predictions using SEPPA

1). The homepage of SEPPA webserver at http://lifecenter.sgst.cn/seppa/index.php was opened.

2). The pdb structure of interest was uploaded. SEPPA also accepts the PDB names and the Chain identifiers. The threshold was left to the default values. Under the default threshold, a sensitivity of 0.568 was received while the specificity was 0.740 on SEPPA training dataset.

Users can set different thresholds. Under a lower threshold, more residues will be included as predicted epitope residues. That always results in the increasement of the true positive rate and the false positive rate at the same time.

Multiple PDB ID entries can be submitted in batch query. Each entry should include PDB ID and chain ID(s), which are separated with space(s) in one line.

3). on submitting the structure of interest, the results mark the epitope residues highlighted in yellow with the core residues shown in lower case. Moreover the 3d structure can also be viewed.
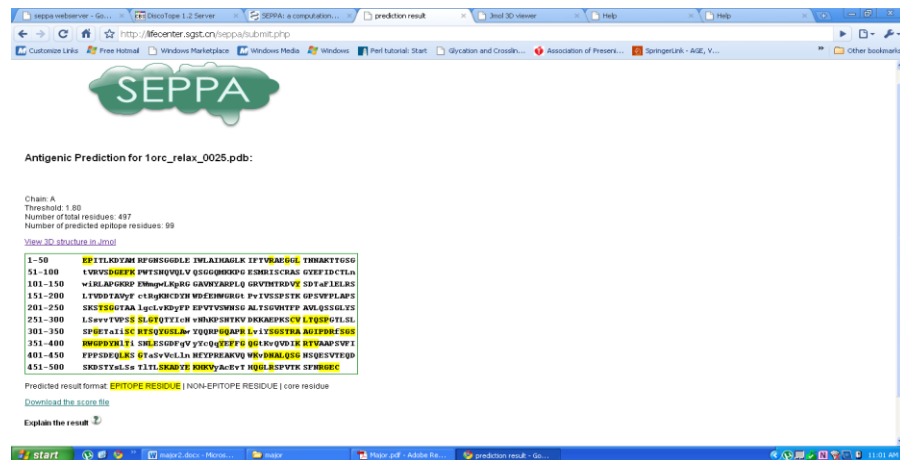
Predictions using PEPITO

1). The homepage of the PEPITO/BEPro webserver at http://pepito.proteomics.ics.uci.edu/index.html was opened.

2). BEPro accepts PDB files i.e the tertiary structure of the antigen so in order to submit, the user must upload a pdb file.

Users can request either a text file with a simplified PDB containing the predicted epitope score in the temperature field (i.e., the last column) or a visualization of the protein using Jmol colored by the epitope score.

Higher scores correspond to a higher likelihood that the residue is in an epitope.

3). The antigen structure was submitted and the results were obtained in form of a pdb file.

(appendix 2)

Prediction using CEP

1). The homepage of CEP at http://bioinfo.ernet.in/cep.htm was opened
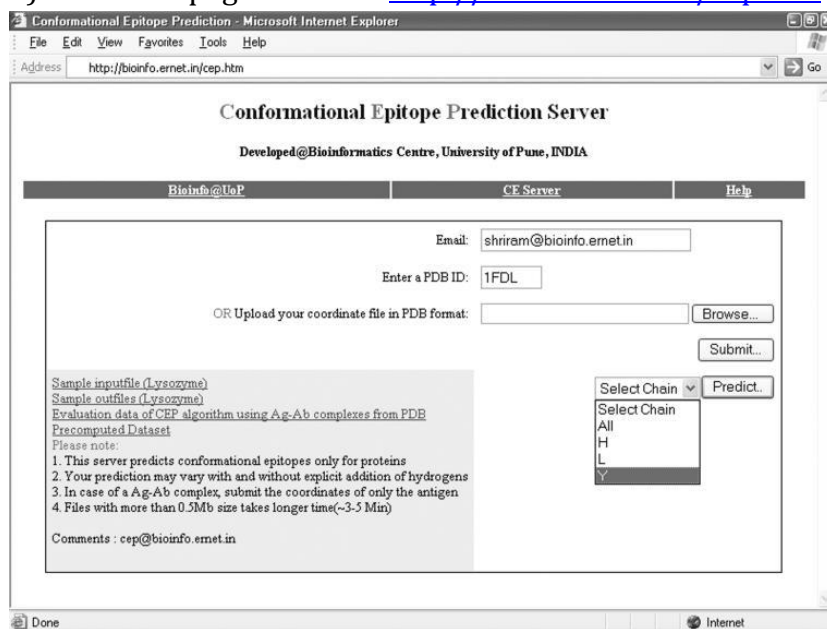
2). CEP server requires the 3D coordinate data in PDB format. The antigen of interest gp120 was submitted by uploading the respective pdb file. The specific chain was specified and the results were mailed at the entered email address. The antigenic determinants, conformational epitopes, and the sequential epitopes were visualized. Also the experimentally determined binding sites were reported.

2. **GENERATION OF A CULLED PDB**

The PDB structures of the candidate scaffolds were downloaded satisfying the below mentioned criteria from the Protein Data Bank and a local database was set up.

Scaffolds selected were:

- Monomeric
- Single chains
- Without ligands
- Having no cysteine residues
- Low molecular weight (<= 40kda)
- Structures having a resolution better than 3A°
- From any source organism except human

A total of 431 pdb structures were downloaded and screened for grafting the epitope regions.

3. <u>**SCAFFOLD SELECTION BY ROSETTA MULTIGRAFTMATCH**</u>

The matching algorithm was developed to select scaffolds suitable for transplantation of structural motifs, based on structural compatibility with a given motif and on their predicted ability to interact with a desired binding partner.

The iterative and exhaustive nature of the matching algorithm, searches all the residues of protein structures in the Protein Data Bank (PDB) for grafting sites suitable for a given structural motif.

The starting input for Multigraft Match is the set of atomic coordinates of the motif in complex with the desired binding partner.

Matches could be anywhere on the protein but to make sure that they are on the surface the clashes must be minimum.

- **<u>Files required</u>**
  1. The PDB files of all the candidate scaffolds. Dssp files are also used though not necessary.
  2. LIST file: This is a text file containing the list of pdb ids of all the candidate scaffolds. The list should contain one pdb per line.
  3. Loop ranges file: This is a text file which contains the information about the number of epitopes to be grafted as well as the residue range of the epitope stretches to be considered for grafting. (appendix 3)
  4. Native complex file: This is a pdb file which contains the coordinates of the antibody and the epitope regions of interest(appendix 3). In this PDB file, Ab must be first! Residues in antigen/epitope section must be in proper numerical order, e.g. 34, 35, 100, 101 .. to N. Residue numbers themselves don't have to

run from 1 to N. All residues specified in each of the 'full ranges' in the loop ranges file must be in the antigen/epitope section.
5. A command file is prepared containing the command.

Command used:
```
rosetta.release -epi_graft -match -rough_match -input_file <name
of the list file> -paths paths.txt -output_file <name of output
file> -loop_ranges <name of loop ranges file> -nres_Ab <number
of residues of the Ab only> -S_align -native_complex <the native
complex file name> -max_closure_rms <default value or any other
as required> -rough_match_closure_rms <default or any other
generally more than max_closure_rms value> -rb_move -
max_intra_clash <default or any other> -max_inter_clash <default
or any other> -no_single_match_output -no_partial_match_output
```

Running Rosetta Epigraft/MultigraftMatch
The program runs by typing "sh <name of the command file>" and then pressing enter.

4. COMPUTATIONAL GRAFTING OF DISCONTINUOUS EPITOPE USING ROSETTA MULTIGRAFTDESIGN

MultigraftDesign takes the multigraft match result as input, removes the appropriate regions of the scaffold, built new segments to connect the epitope to the scaffold and designed the side chains neighboring the epitope and connecting segments to support the graft.

This involved aggressive structural manipulations, including replacements of ordered secondary structure motifs by the epitope segments and sequence design of 10 or more core residues.

MultigraftDesign was used for backbone grafting which alters both the sequence and structure of the scaffold to accommodate a desired functional motif.

Several design variants of each candidate scaffold were generated.

- **Files required**
  1. The PDB files of all the candidate scaffolds. Dssp files are also used though not necessary.
  2. LIST file: This is a text file containing the list of pdb ids of all the candidate scaffolds. The list should contain one pdb per line

3. Loop ranges file: This is a text file which contains the information about the number of epitopes to be grafted as well as the residue range of the epitope stretches to be considered for grafting. (appendix 3)
4. Native complex file: This is a pdb file which contains the coordinates of the antibody and the epitope regions of interest. In this PDB file, Ab must be first! Residues in antigen/epitope section must be in proper numerical order, e.g. 34, 35, 100, 101 .. to N. Residue numbers themselves don't have to run from 1 to N. All residues specified in each of the 'full ranges' in the loop ranges file must be in the antigen/epitope section.

5. Input file: Typically this file contains match results that come from the Epi-Graft Match run but with removal of all data columns AFTER the last translation entry 'T_z'. All of the data columns such as 'overall_rms' cannot be in the input file! Additional columns are included into the input file which indicates the type of grafting to be performed on the epitope.(appendix 3)

Example input file:

(match result columns) $ graft_bb graft_sc scaffold_tether_nterm scaffold_linker_nterm scaffold_moveable_nterm |scaffold_tether_cterm scaffold_linker_cterm scaffold_moveable_cterm | epitope_tether_nterm epitope_linker_nterm epitope_moveable_nterm | epitope_tether_cterm epitope_linker_cterm epitope_moveable_cterm perturb_rb? move_all_epitope_residues? close as_single_break? repack_epitope_residues?
(my_scaffold.pdb   362 379 ...) $ true   true   3 -1 3 | 4 2 2 | 0 0 0 | 0 0

0

(my_scaffold.pdb   478 482 ...) $ true   true   4 0 4 | 4 0 4 | 0 3 3 | 1 0

1

Interpretation of the columns in the input file

1. The '$' symbol is REQUIRED to separate the match result columns and the graft info columns, since the number of match result columns and graft info columns can be variable.
2. 'graft_bb' column is 'true' or 'false' and indicates whether or not to graft the backbone of the corresponding epitope. 'graft_sc' column 'true' or 'false' and indicates whether or not to graft the sidechain of the corresponding epitope. 'graft_bb' takes precedence over 'graft_sc', so if 'graft_bb' is true, then 'graft_sc' is forced to true in the software. If 'graft_bb' is false then no other columns are necessary.

3. The latter 15 columns of each line define the number of n-terminal and c-terminal tether/linker/moveable residues on the scaffold and epitope sides of each epitope loop.
4. A total of three separate '|' must be used to separate the entry groups.
5. Tether residues are the total number of residues to one side of the *-terminus, linker residues indicate the number of residues to grow/remove from the scaffold side of the chainbreak, and moveable residues are the total number of moveable residues during loop closure.
6. There are restrictions on the values of the tether/linker/moveable residues:

Restriction: 'tether_*term' >= 0

Restriction: 'tether_*term' + 'linker_*term' >= 0

Restriction: 'tether_*term' + 'linker_*term' >= 'moveable_*term'

Restriction: epitope_tether_nterm < 1 + (length of epitope loop)/2

Restriction: epitope_tether_cterm < (length of epitope loop)/2

Restriction: epitope_linker_*term >= 0

7. The next (optional) 2 columns of each line indicate whether that epitope component should be rigid body perturbed prior to each build loops attempt, and whether or not to move all residues in the epitope when attempting to close.
8. The next (optional) column of each line indicates whether or not the epitope component should be grafted using single break closure, rather than double break closure.
   Currently this requires move_all_epitope_residues to be true. Note that this removes the rigid body dependence of the epitope components! This column is ignored for primary components. Single break closure and rigid body perturbation cannot both be true at the same time.
9. The last (optional) column of each line indicates whether or not to forcibly repack the epitope residues. This is recommended if you end up moving any epitope residues, like with "move_all_epitope_residues". This column is valid for ALL components.

Command used

```
rosetta.release  -epi_graft  -multigraft  -nres_Ab  <number  of
residues of the Ab> -native_complex <name of native complex pdb
file> -loop_ranges <name of loop ranges file> -input_file <input
file  name>  -use_non_monotone_line_search  -ex1  -ex1aro  -ex2  -
```

```
extrachi_cutoff  0  -atom_vdw_set  highres  -close_as_ALA  -
grow_as_ALA -graft_with_Ab -repack_Ab -build_loops -refine_loops
-design_after_closure      -refine_with_Ab_after_design      -
closure_attempts    <number    of    closure    attempts>    -
store_n_best_closures <number of closures to be stored> -
dump_all_closure_attempt_structures -design_attempts <number of
design attempts> -store_n_best_designs <number of designs to be
stored> -refine_with_constraints -intra_design_cutoff <default
or any other as required> -inter_design_cutoff <default or any
other as required> -max_chainbreak_score <default or any other>
-max_local_rama <default or any other> -design_after_closure -
Ab_epitope_optimize -repack_epitope -use_old_graft_info_format -
paths paths.txt -output_file multi_nw.out -vall <name of vall
database> -export_resfile -export_blueprint
```

Running Rosetta MultigraftDesign

The program runs by typing "sh <name of the command file>" and then pressing enter.

6. ENERGY MINIMIZATION BY ROSETTA FASTRELAX

After grafting of the discontinuous epitopes on the scaffolds, these structures were minimized.

Relax application in Rosetta carries out the task of simple structural refinement of the full atom Rosetta models. It can also read the centroid models in which case it will convert the model into a full atom model and pack the sidechains.

It takes as input one or more structures in PDB format.

It is a general purpose protocol and in most cases relax is the last step in a larger protocol.

The lowest energy structures are of interest to the user.

The structure can change upto 2-3A from the starting conformation.

The structures obtained after design experiment were relaxed using this application

25 structures were generated for each selected design (scaffold with the epitope).

The lowest energy structures were selected for each designed structure.

Files required

1. PDB file of the designed structure (containing the Ab-scaffold with the epitopes) to be minimized
2. Commands file containing the command to be used.

   Command used
   ```
   /root/rosetta_source/bin/relax.default.linuxgccrelease
   @<script file name> -run:constant_seed -nodelay -database
   <path to the rosetta database> -overwrite
   ```

3. Script file containing the script (appendix3)

```
-s <pdb file name>
-constant_seed
-out:nstruct <number of structures to be generated in the
output>
-out:output
-out:pdb
-out:show_accessed_options
-relax:quick
-mute core.util.prof ## dont show timing info
-unmute core.io.database
-out:file:fullatom
-out:path:pdb
```

Running RosettaFastRelax

The program runs by typing "sh <name of the command file>" and then pressing enter.

7. VALIDATION

- 4 best structures were selected after the energy minimization and were hierarchically arranged as per the following evaluation criteria:

- **Rmsd**: the rmsds between the designs and the actual structure (Gp120-VRCO1) were determined using pymol.
- **Computational alanine scanning** was performed for each (scaffold with the epitope+VRCO1).
- **Interactions** of the actual structure (gp121-VRCO1) and the designs (scaffold with the epitope-VRCO1) were determined and compared. This was done using PIC server.

73

RESULTS

*actual pdb ids and some data has been concealed from the report due to confidentiality issues of this work at I.C.G.E.B NewDelhi. I do not hold the permission to reveal them.

1. ## Results of computational alanine scanning (for defining the epitopes)

   Computational alanine scanning was performed for the complex G.pdb (VRCO1-gp120 complex) to determine the hotspot residues and confirm that the hotspots are determined within the LoopD and CD4bs.

   Gp120 comprises of four identical chains, so only the G chain of gp120 was considered for scanning the interface between the G and H (heavy chain of VRCO1) and G and L (light chain of VRCO1). It was found that most of the hotspots lie within the LoopD (276-284) and CD4bs (364-371) regions implying their important interactions with the VRCO1.

pdb#	chain	int_id	res#	aa	DDG(complex)	DDG(complex,obs)	DG(partner)
275	G	1	159	4	0.09	0	-0.06
276	G	1	160	12	0.15	0	-0.22
278	G	1	162	17	0.11	0	0.05
279	G	1	163	12	2.02	0	1.11
280	G	1	164	12	2.76	0	-0.33
282	G	1	166	9	0.56	0	0.69
365	G	1	226	16	0.55	0	-0.09
368	G	1	229	3	1.55	0	-0.94
371	G	1	232	8	1.06	0	1.12
425	G	0	277	12	-0.02	0	0.63
455	G	1	307	17	0.29	0	0.7
457	G	1	309	3	-0.14	0	0.96
461	G	1	313	12	0.9	0	-0.8
462	G	1	314	12	0.91	0	-0.02
469	G	1	321	15	0.21	0	4.73
474	G	0	326	12	-0.02	0	0.92
476	G	1	328	9	-0.11	0	0.85
3	L	1	347	18	0	0	0.7
91	L	1	433	10	1.17	0	0.36
96	L	1	434	4	0.8	0	-0.02
97	L	1	435	5	0.59	0	0.73
47	H	1	600	19	0.6	0	5.96
50	H	1	603	19	3.38	0	3.79
52	H	1	605	9	0.37	0	0.75
52	H	0	607	18	0.21	0	0.25
53	H	1	608	17	0.83	0	0.83
56	H	0	611	18	0.6	0	1.54
57	H	1	612	12	4.08	0	0.21
58	H	0	613	5	0.15	0	3.92
60	H	0	615	16	0.16	0	-0.14
64	H	1	619	15	2.01	0	0.56
69	H	0	624	10	0	0	3.69
71	H	1	626	15	1.56	0	2.46
73	H	1	628	15	1.24	0	0.65
74	H	1	629	3	-0.07	0	-1.04
98	H	1	656	20	-0.13	0	0.09
99	H	1	657	17	0.26	0	1.02
100	H	1	662	19	2.98	0	3.54

Interpretation of the output: the columns contain the following information

pdb#, number of mutated residue in the pdb file.

chain, pdb chain identifier.

int_id: measure of whether a residue side chain atom is within 4 Å of an atom on the other partner (1) or not contacting directly, but buried upon binding (0).

res#, continuous residue numbering of all partners, starting with residue number 1.

aa, amino acid type according to one-residue nomenclature in alphabetical order (1-A, 2-C, 3-D, 4-E ...)

ΔΔG(complex), predicted change in binding free energy upon alanine mutation.

ΔΔG(complex, obs), observed changes in binding free energy upon alanine mutation (user input in mutation list, otherwise set to zero)

ΔG(partner), predicted change in protein stability of the mutated complex partner upon alanine mutation

Hot-spot residues can be defined operationally as those for which alanine mutations have destabilizing effects on ΔΔGcomplex of more than 1 kcal/mol. For a comparison with experimental data, a correctly identified hot spot means a residue with a predicted and observed ΔΔGcomplex value of greater than or equal to 1 kcal/mol, and a correctly identified neutral residue has both predicted and observed ΔΔGbind values of less than 1 kcal/mol.

2. **Results of epitope prediction**

   Different servers were used for discontinuous Bcell epitope prediction and the results from all the servers were used to generate a consensus. It was found that all the servers could predict the epitopic residues within the regions of LoopD (276-284) and CD4bs (364-371) (result table in appendix 1 and appendix 2).
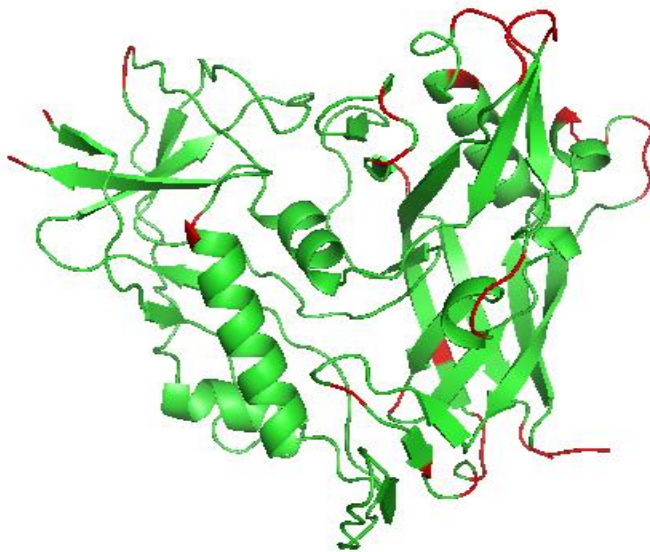

Figure 11: gp120 with the predicted discontinuous epitopes (shown in red)

3. **Designed Structures (scaffolds carrying the grafts)**

a. **A.pdb:** A.pdb is a crystal structure of single chain monomeric protein of 71 amino acid residues and the source organism is Enterobacteria phage lambda. It was used as a candidate scaffold and the loopD and CD4bs were successfully grafted onto the scaffold.

Sequence of the designed immunogen: red regions are the LoopD and CD4bs respectively.
EPITLKDYAMRFGNSGGDLEIWLAIHAGLKIFTVRAEGGLTNNAKTTGSGTVRVSDGE FKPWTSN
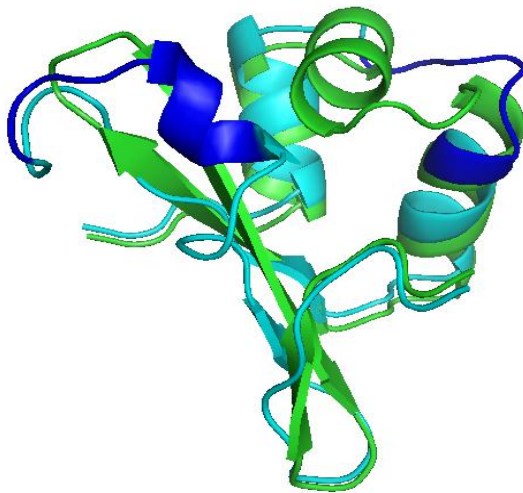**Structure of the designed immunogen:**

Figure 12: alignment of the actual scaffold (green) with the modified scaffold (cyan). In the modified scaffold it is shown that new region of the epitopes LoopD and CD4bs (shown here in dark blue) are constructed by replacing the backbone of the scaffold.

b. **C.pdb:** It is a refined crystal structure from staphylococcus aureus protein. It is a single chain monomeric protein of 257amino acid long. Two different designs were constructed using this scaffold.

Sequence of the designed immunogen: red regions are the CD4bs and loopD respectively.

1. KELNDLEKKYNAHIGVYALDTKSGKEVKFNSDKRFHHASTSFAIWSA
   IVLEQVPYNKLNKKVHINKSDAHDYSPILEKYVGKDITLKALIEAAMT
   YSDDTAAHELMAQGSENLTNNAKTAGLGAQLVKIRLKELGDKVTSPG
   TPPSGHTSGPSGGDLEITGNGPDGSTPAAFGKTLNKLIANGKLSKENK
   KFLLDLMLNNKSGDTLIKDGVPKDYKVADKSGQAQVYASRNDVAFVY
   PKGQSEPIVLVIFTSKDNKSDKPNDKLISETAKSVMKEF

2. KELNDLEKKYNAHIGVYALDTKSGKEVKFNSDKPFAPASTSDAIWSAILLEYVPYNKKVHINKSD
   AVDYSPILEKYVGKDITLKALIEASMTYSDNTASIKIMEVGALLTNNAKTAGGASNGVKNRLKELG
   DKVFDPTTTFHDLSSGGPSGGDLEIRGLPQQGGPTPASVGKTLNKLIANGKLSKENKKFLLDLML
   NNKSGDTLIKGVPKDYKVADKSGQAITYASRNDVAFVYPKGQSEPIVLVIFTNKDNKSDKPNDKLI
   SETAKSVMKEF
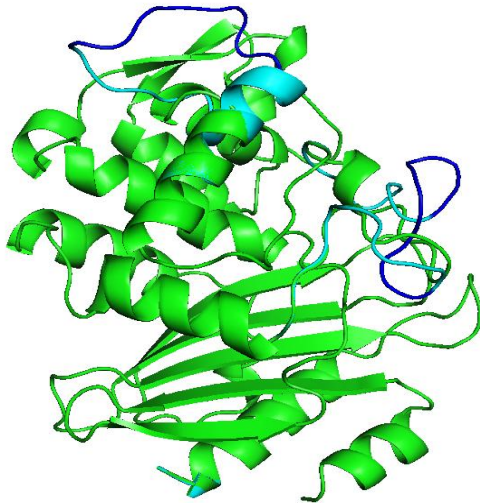
Structure of the designed immunogen:

Figure 13: alignment of the actual scaffold (green) with the modified scaffold (cyan). In the modified scaffold it is shown that new region of the epitopes LoopD and CD4bs (shown here in dark blue) are constructed by replacing the backbone of the scaffold.

c. B.pdb: B.pdb is a crystal structure of a hypothetical protein from Pyrococcus horikoshii Ot3. It is a single chain monomeric protein of length 238 amino acid.

Sequence of the designed immunogens: red regions are the loopD and CD4bs respectively.
MKEEAVRVIEEVLKQGRTAMVEYEAKQVLKAYGLPVPEEKLAKTLDE
ALEYAKEIGYPVVLKLMSPQILHKSDAKVVMLNIKNEEELKKKWEEI
HENAKKYRPDAEILGVLVAPMLKPGRAVIIRTGPAGKSGGDLEEGGG
AIIFFHLGPFQWMQALTNNAKTSTDYFPFIPITEKRARYMIQSIEAYPI

LAGAEEPADIDAIVDMLLKVSKLMYDLEDYITTSELNPVFVYNKGEGA
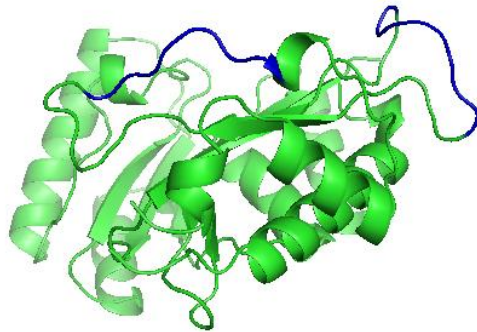VIVDSSITLFPK.

Structure of the designed immunogen

Figure 14: In the modified scaffold it is shown that new region of the epitopes LoopD and CD4bs (shown here in dark blue) are constructed by replacing the backbone of the scaffold.

4. **Validation of the designed structures**

   a. The designed structures were validated by determining their rmsd with the actual complex G.pdb (gp120-VRCO1 complex). Further their interactions at the interface were determined and compared with the actual one (G.pdb). The designs with maximum interactions identical to the G.pdb were determined using PIC webserver (appendix 3) and were arranged hierarchically. Also computational alanine scanning was again performed for all the design-Ab complexes.

   Out of the four scaffolds, the immunogen designed using C.pdb showed maximum interactions similar to the actual complex C.pdb (gp120-VRCO1)

Interactions at the interface of G.pdb (gp120-VRCO1) complex	Interactions at the interface of designed scaffold C.pdb and VRCO1 complex

Ala(281)→hydrophobic interaction→trp(50)(H) Ala(281)→hydrophobic interaction→trp(100)(H) Ile(371)→hydrophobic interaction→ala(56)(H) Asp(368)→main chain interaction→gly(54)(H) Lys(282)→mainchain-sidechain→asp(99)(H) Asp(368)→mainchain-sidechain→gly(54)(H) Thr(278)→sidechain interactions→tyr(91)(L) Asn(280)→sidechain interactions→glu(96)(L) Asp(368)→ionic interactions→arg(71)(H)	Ala(117)→hydrophobic interaction→trp(100)(H) Ala(117)→hydrophobic interaction→tyr(91)(L) Asp(156)→main chain interaction→gly(54)(L) Thr(114)→mainchain-sidechain→tyr(28)(L) Asp(156)→mainchain-sidechain→gly(54)(H) Thr(114)→sidechain interactions→tyr(91)(L) Asn(115)→sidechain interactions→gln(27)(L) Asp(156)→ionic interaction→arg(71)(H)

b. The rmsd of the designed structures with the actual complex was also determined.

1. Scaffold B.pdb

Figure 15: Fig: the image shows the alignment of the designed scaffold-Ab complex (B.pdb-VRCO1) with the actual complex G.pdb (gp120-VRCO1). The red regions are the epitopes on gp120 and the dark blue regions are the epitopes grafted on the scaffold (green). This clearly shows that the two epitope loops are grafted (dark blue) at almost the same relative distance from each other as present in the actual complex (red). The rmsd was found to be 1.236A°
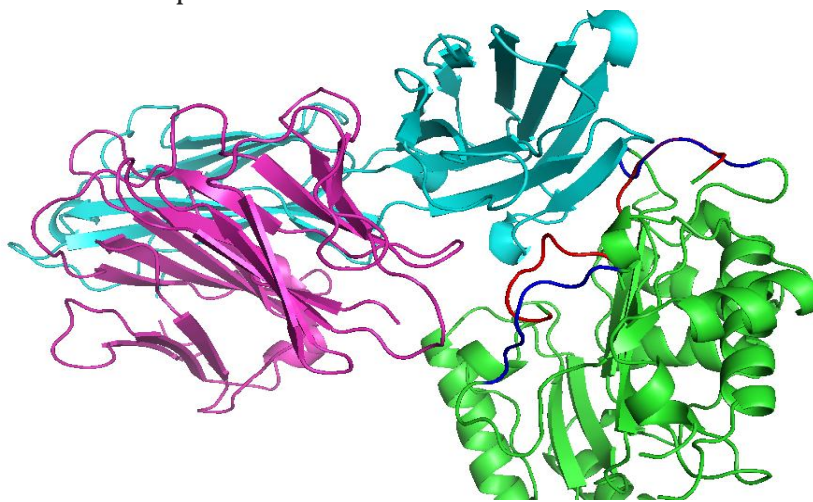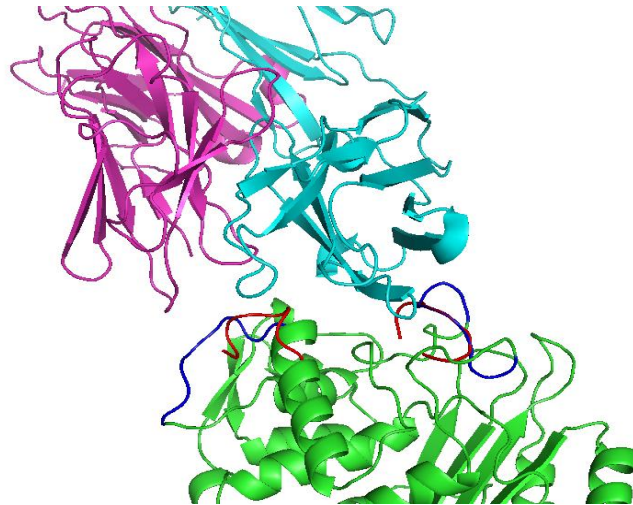

2. Scaffold C.pdb

Figure 16: the image shows the alignment of the designed scaffold-Ab complex (C.pdb-VRCO1) with the actual complex G.pdb (gp120-VRCO1). The red regions are the epitopes on gp120 and the dark blue regions are the epitopes grafted on the scaffold (green). This clearly shows that the two epitope loops are grafted (dark blue) at almost the same relative distance from each other as present in the actual complex (red). The rmsd was found to be 1.707A° and 2.956A° for both the designs respectively.

3. Scaffold A.pdb
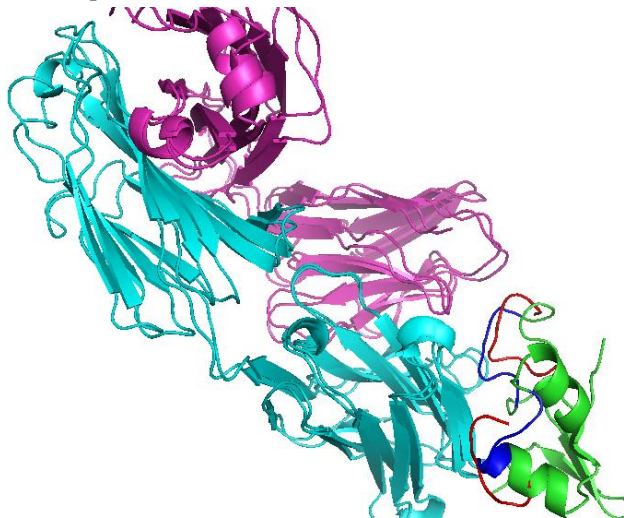
Figure 17: the image shows the alignment of the designed scaffold-Ab complex (A.pdb-VRCO1) with the actual complex G.pdb (gp120-VRCO1). The red regions are the epitopes on gp120 and the dark blue regions are the epitopes grafted on the scaffold (green). This clearly shows that the two epitope loops are grafted (dark blue) at almost the same relative distance from each other as present in the actual complex (red). The rmsd was found to be 2.347A°

CONCLUSION

The designing of HIV immunogens by grafting of discontinuous Bcell epitopes on suitable scaffolds is a step towards a new generation of epitope based vaccines. This has become an important and an intelligent strategy since the discovery of broadly neutralizing antibodies that target certain conserved regions on the surface glycoprotein gp120. These broadly neutralizing antibodies are highly potent and are effective against a wide range of HIV viruses.

In the present work, HIV immunogen designing was performed by grafting the discontinuous Bcell epitopes (on the gp120) on certain carrier proteins called scaffolds. The scaffolds selected were low molecular weight, monomeric, single chain proteins, without disulphide linkages, having a good resolution and were from any source organism excepting humans. Backbone grafting was performed as this type of grafting is very useful in cases where no structural similarity exists between the scaffold and the epitope regions to be grafted. Moreover it has been observed that backbone grafting generated immunogens with a greater binding affinity for the antibody than the side chain grafting. Hence backbone grafting gives a completely new backbone structure as well as an entirely new sequence. Important consideration while performing grafting of the epitope is that the grafts are made on the surface and are in the same conformation as present on the actual

antigen. To make sure that the grafts come on the surface the clashes (i.e. both the inter and intra clash values) had to be minimum. Generally a value of 50 to 100 gives biologically significant and surface exposed grafts. The designed complexes were minimized to attain the lowest energy conformations and the most stabilized structures having the lowest energy are of interest to the user.

Validation of the designs requires analysis of the rmsds of the designed imunogen-Ab complex with the original complex (gp120-VRCO1 complex), the comparision of interactions at the interface and the determination of the important residues at the interface of the designed immunogen-Ab complex by computational alanine scanning and hence comparing them with the original complex. Based on the above analysis the four best designs were arranged hierarchically as follows:

1. C_1.pdb
    DAKLMLEKKYNAHISGYTILLKTGKEVKFNSDHTYDDSTTSFAIWSA
    IVLEQVPYINTKKYLNGNKSDAHDYSPILEKYVMTAILNTPLIEAAMT
    YSDDTAAHELMAQGSENLTNNAKTAGLGAQLLTNNGLEIPSGGTSPG
    TGYSGHTSGPSGGDLEITGNGPDGSTPAWWRRSTYLIANGKLSKENK
    KFLLDLMLNNKSGDTLIKDGVPKDYKVADKSGQAQVYRDYTASAFVY
    PKGQSEPIVLVIFNTPDNKSNNPGYHLISETAKSVMKEF

2. B.pdb
    MNTFGVDDSTAGGPYTHHANTKYEAKQVLKAYGLPVPEEKLAKTLD
    EALEFLGDDHPVVLKLMSRQSLHKSDANTLIGDAVSNEEELKKKWEE
    IKTNNGKYRPDAEILGVVRADGSTVNGLIVIIRTGPAGKSGGDLEEGGG
    LAMNITLGPFQWMQALTNNAKTSTDYFPFIPITEKRAMSPIQSIEAYPI
    LAGAEEPSPRAHALLKVSKLMYDLDINESHSIRVESPAPVFVYNKGEGA
    GATEWAYSITLFPK.

3. A.pdb
    EPRAMANSMRFGNSGGDLEIWLAIWAYLKIGATEGGLTNNAKTTGSGSASVSDGNSGPWLTD

4. C_2.pdb
    ABHISHEKKYNAHGARYADIVYAGKEVKFNSDKPFAPASTSDAIWSMILEYPVPYNKKVHNDASD
    DVDYSPILSAARAAHITLKALIEASMTYSPIDERMANMEVGALLTNNAKTAGGASNGNEVIAKEDG
    DKVFDPTTTFHDLSSGGPSGGDLEIRGLPQQGGATVSSVGKPLAYERANGMNSKENKKVYLGDLML
    NKDSGDTLIKGSTVYAYKVADMNPGQAITLGSRNDVCDVYPKGFEASTVLVIFTNDSTPLDKPNDKLI
    SFELLAWMKEF

DISCUSSION

With 90% of the Bcell epitopes being discontinuous, the designing of HIV immunogen by computational grafting of discontinuous epitopes on appropriate scaffold proteins to make them capable of eliciting the broadly neutralizing antibodies has challenges of its own.

First and foremost the reliable prediction of highly immunogenic epitopes to significantly reduce experimental efforts in discovering epitopes needed for the designing of vaccines and immunodiagnostics is a trivial task and requires the knowledge of protein structures. This is more difficult in case of discontinuous Bcell epitopes where the epitopes are entirely context dependent and hence a binary classification is simply not reliable. So to make our predictions more reliable combinations of different servers must be considered. When the aim of the immunological study is to identify as many binders as possible, servers should be used in union mode and when efficiency is a priority, and experimental work aims to pick out only a few, highly probable binders, server outputs should be combined using the intersection mode.

Understanding the mechanism of action of the broadly neutralizing antibodies is required to reveal the vulnerabilities that are shared by many different types of HIV. These conserved regions form the basis of epitope driven vaccines. Of particular importance are the CD4bs targeting neutralizing antibodies because it is this site which is present in all types of HIV to gain access to the host cell CD4 receptors. Recently discovered VRCO1 is

important because it is highly potent and manages to evade the glycan shield of the HIV. The antibody does this by utilizing the glycans to gain access to the CD4 binding site and hence rather than being obstructed, it uses the glycans to target the CD4 binding site. With such potent broadly neutralizing antibodies being discovered, interest in the field of HIV vaccine research has been reignited.

Secondly the grafting of the discontinuous epitope in exactly the same conformation and same relative orientation with respect to each other as in the original gp120 structure is difficult to achieve and requires a thorough study of the actual structure. The more similar the designs are to the actual epitopes on the antigens, more closely the interactions mimic the actual ones. While grafting, intelligent selection of the number of linkers to be added ensures that the structures are properly closed and are flexible enough to have stable closures.

Thirdly, selection of highly immunogenic scaffolds of carrier proteins must be carefully done because it ia a possibility that the immunogenicity due to the scaffold takes over the immunogenicity due to the transplanted motif. So the interface has to be carefully designed so that the epitope forms maximum interactions with the neutralizing antibody.

Although HIV vaccine research continues to lure us, we must not neglect the fact that HIV virus requires not only a strong humoral response and the broadly neutralizing antibodies but also a strong cell mediated immunity as well as the targeted action of many non-neutralizing antibodies. Also a major problem with the virus is the long inactive period in many patients, making its detection and diagnosis even more difficult. Hence novel strategies are being discovered for rational designing of HIV vaccine.

REFERENCES

1. Piot P, Quinn TC, Taelman H, Feinsod FM, Minlangu KB, Wobin O, Mbendi N, Mazebo P, Ndangi K, Stevens W, et al.1984. Acquired immunodeficiency syndrome in a heterosexual population in Zaire. Lancet 2: 65–69.
2. Van de Perre P, Rouvroy D, Lepage P, Bogaerts J, Kestelyn P, Kayihigi J, Hekker AC, Butzler JP, Clumeck NA. 1984. Acquired immunodeficiency syndrome in Rwanda. Lancet 2: 62–65.
3. Serwadda D, Mugerwa RD, Sewankambo NK, Lwegaba A, Carswell JW, Kirya GB, Bayley AC, Downing RG, Tedder RS, Clayden SA, et al. 1985. Slim disease: A new disease in Uganda and its associationwith HTLV-III infection. Lancet 2: 849–852.
4. Gouws E, Abdool Karim Q. 2010. HIV infection in South Africa: The evolving epidemic. In HIV/AIDS in South Africa, 2nd ed., pp. 55–73. Cambridge University Press, Cambridge.
5. PAHO, WHO, UNAIDS. 2001. Monitoring the AIDS Epidemic (MAP). HIV and AIDS in the Americas: An epidemic with many faces. http://www.who.int/hiv/strategic/pubrio00/en/index.html.
6. Phanuphak P, Locharernkul C, Panmuong W, Wilde H.1985. A report of three cases of AIDS in Thailand. Asian Pac J Allergy Immunol 3: 195–199.
7. Ma Y, Li Z, Zhao S. 1990. HIV infected people were first identified in intravenous drug users in China. Chin J Epidemiol11: 184–185.
8. Estebanez P, Fitch K, Najera R. 1993. HIV and female sex workers. Bull World Health Organ 71: 397–412.
9. Dore GJ, Kaldor JM, Ungchusak K, Mertens TE. 1996. Epidemiology of HIV and AIDS in the Asia-Pacific region. Med J Aust 165: 494–498.
10. Crofts N, Reid G, Deany P. 1998. Injecting drug use and HIV infection in Asia. AIDS 12: 69–78.
11. WHO. 1998. Fifty years of WHO in the Western Pacific Region, Chapter 26. Sexually transmitted diseases, including HIV/AIDS, pp. 245–259. http://www.wpro.who.int/NR/rdonlyres/3E44D634-C0F9-480C-B6C2-1C418AE8F3A0/0/chapter26.pdf.
12. Los Alamos HIV Sequence Database. 2010. Overview of the subtypes of primate immunodeficiency viruses. http://www.hiv.lanl.gov/content/sequence/HelpDocs/subtypes.html.
13. Hemelaar J, Gouws E, Ghys PD, Osmanov S. 2006. Global and regional distribution of HIV-1 genetic subtypes and recombinants in 2004. AIDS 20: W13–W23.
14. Schultz AM, Bradac JA. 2001. The HIV vaccine pipeline, from preclinical to phase III. *AIDS* 15(Suppl. 5):S147–58
15. Moore JP, Cao Y, Qing L, Sattentau QJ, Pyati J, et al. 1995. Primary isolates of human Immune deficiency virus type 1 are relatively resistant to neutralization by monoclonal antibodies to gp120, and their neutralization is not predicted by studies with monomeric gp120. *J. Virol.* 69:101–9.

16. Sattentau QJ, Moore JP. 1995. Human immunodeficiency virus type 1 neutralization is determined by epitope exposure on the gp120 oligomer. *J. Exp. Med.* 182:185–96
17. Fouts TR, Binley JM, Trkola A, Robinson JE, Moore JP. 1997. Neutralization of the human immunodeficiency virus type 1 primary isolate JR-FL by human monoclonal antibodies correlates with antibody binding to the oligomeric form of the envelope glycoprotein complex. *J. Virol.* 71:2779–85
18. Parren PWHI, Mondor I, Naniche D, Ditzel HJ, Klasse PJ, et al. 1998. Neutralization of human immunodeficiency virus type 1 by antibody to gp120 is determined primarily by occupancy of sites on the virion irrespective of epitope specificity. *J. Virol.* 72:3512–19
19. Schonning K, Bolmstedt A, Novotny J, Lund OS, Olofsson S, Hansen JE. 1998. Induction of antibodies against epitopes inaccessible on the HIV type 1 envelope oligomer by immunization with recombinant monomeric glycoprotein 120. *AIDS Res. Hum. Retrovir.* 14:1451–56
20. Poignard P, Moulard M, Golez E, Vivona V, Franti M, et al. 2003. Heterogeneity of envelope molecules expressed on primary human immunodeficiency virus type 1 particles as probed by the binding of neutralizing and nonneutralizing antibodies. *J. Virol.* 77:353–65
21. Herrera C, Spenlehauer C, Fung MS, Burton DR, Beddows S, Moore JP. 2003. Nonneutralizing antibodies to the CD4-binding site on the gp120 subunit of human immunodeficiency virus type 1 do not interfere with the activity of a neutralizing antibody against the same site. *J. Virol.* 77:1084–91.
22. Wyatt R, Sodroski J. 1998. The HIV-1 envelope glycoproteins: fusogens, antigens, and immunogens. *Science* 280:1884–88.
23. Fenouillet E, Gluckman JC, Jones IM. 1994. Functions of HIV envelope glycans. *Trends Biochem. Sci.* 19:65–70
24. Leonard CK, Spellman MW, Riddle L, Harris RJ, Thomas JN, Gregory TJ. 1990. Assignment of intrachain disulfide bonds and characterization of potential glycosylation sites of the type 1 recombinant human immunodeficiency virus envelope glycoprotein (gp120) expressed in Chinese hamster ovary cells. *J. Biol. Chem.* 265:10373–82
25. Parren PWHI, Burton DR, Sattentau QJ. 1997. HIV-1 antibody–debris or virion? *Nat.Med.* 3:366–67
26. Burton DR, Montefiori DC. 1997. The antibody response in HIV-1 infection. *AIDS* 11(Suppl. A):S87–98.
27. Modrow S, Hahn BH, Shaw GM, Gallo RC, Wong-Staal F, Wolf H. 1987. Computer assisted analysis of envelope protein sequences of seven human immunodeficiency virus isolates: prediction of antigenic epitopes in conserved and variable regions. *J. Virol.* 61:570–78
28. Willey RL, Rutledge RA, Dias S, Folks T, Theodore T, et al. 1986. Identification of conserved and divergent domains within the envelope gene of the acquired immunodeficiency syndrome retrovirus. *Proc. Natl. Acad. Sci. USA* 83:5038–42
29. Helseth E, Olshevsky U, Furman C, Sodroski J. 1991. Human immunodeficiency virus type 1 gp120 envelope glycoprotein regions important for association with the gp41 transmembrane glycoprotein. *J. Virol.* 65:2119–23
30. Moore JP, Sattentau QJ, Wyatt R, Sodroski J. 1994. Probing the structure of the human immunodeficiency virus surface glycoprotein gp120 with a panel of monoclonal antibodies. *J. Virol.* 68:469–84
31. Pollard SR, Rosa MD, Rosa JJ, Wiley DC. 1992. Truncated variants of gp120 bind CD4 with high affinity and suggest a minimum CD4 binding region. *EMBO J.* 11:585–91
32. Moore JP, Sodroski J. 1996. Antibody cross-competition analysis of the human immunodeficiency virus type 1 gp120 exterior envelope glycoprotein. *J. Virol.* 70:1863–72
33. Olshevsky U, Helseth E, Furman C, Li J, Haseltine W, Sodroski J. 1990. Identification of individual human immunodeficiency virus type 1 gp120 amino acids important for CD4 receptor binding. *J. Virol.* 64:5701–7
34. Moore JP, Willey RL, Lewis GK, Robinson J, Sodroski J. 1994. Immunological evidence for interactions between the first, second, and fifth conserved domains of the gp120 surface glycoprotein of human immunodeficiency virus type 1. *J. Virol.* 68:6836–47
35. Sullivan N, Sun Y, Sattentau Q, Thali M, Wu D, et al. 1998. CD4-induced conformational changes in the human immunodeficiency virus type 1 gp120 glycoprotein: consequences for virus entry and neutralization. *J. Virol.* 72:4694–703

36. Wyatt R, Moore J, Accola M, Desjardin E, Robinson J, Sodroski J. 1995. Involvement of the V1/V2 variable loop structure in the exposure of human immunodeficiency virus type 1 gp120 epitopes induced by receptor binding. *J. Virol.* 69:5723–33

37. Wyatt R, Sullivan N, Thali M, Repke H, Ho D, et al. 1993. Functional and immunologic characterization of human immunodeficiency virus type 1 envelope glycoproteins containing deletions of the major variable regions. *J. Virol.* 67:4557–65

38. Cao J, Sullivan N, Desjardin E, Parolin C, Robinson J, et al. 1997. Replication and neutralization of human immunodeficiency virus type 1 lacking the V1 and V2 variable loops of the gp120 envelope glycoprotein. *J. Virol.* 71:9808–12

39. Ren XP, Sodroski J, Yang XZ. 2005. An unrelated monoclonal antibody neutralizes human immunodeficiency virus type 1 by binding to an artificial epitope engineered in a functionally neutral region of the viral envelope glycoproteins. *J. Virol.* 79:5616–24

40. Rizzuto CD, Wyatt R, Hernandez-Ramos N, Sun Y, Kwong PD, et al. 1998. A conserved HIV gp120 glycoprotein structure involved in chemokine receptor binding. *Science* 280:1949–53

41. Rizzuto C, Sodroski J. 2000. Fine definition of a conserved CCR5-binding region on the human immunodeficiency virus type 1 glycoprotein 120. *AIDS Res. Hum. Retrovir.* 16:741–49

42. Chen B, Vogan EM, Gong H, Skehel JJ, Wiley DC, Harrison SC. 2005. Structure of an unliganded simian immunodeficiency virus gp120 core. *Nature* 433:834–41

43. Kwong PD, Wyatt R, Robinson J, Sweet RW, Sodroski J, HendricksonWA. 1998. Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor
and a neutralizing human antibody. *Nature* 393:648–59

44. Kwong PD, Wyatt R, Sattentau QJ, Sodroski J, HendricksonWA. 2000. Oligomeric modeling and electrostatic analysis of the gp120 envelope glycoprotein of human
immunodeficiency virus. *J. Virol.* 74:1961–72

45. Cheng-Mayer C, Brown A, Harouse J, Luciw PA, Mayer AJ. 1999. Selection for neutralization resistance of the simian/human immunodeficiency virus SHIVSF33A variant in vivo by virtue of sequence changes in the extracellular envelope glycoprotein that modify N-linked glycosylation. *J. Virol.* 73:5294–300

46. Wei XP, Decker JM, Wang SY, Hui HX, Kappes JC, et al. 2003. Antibody neutralization and escape by HIV-1. *Nature* 422:307–12

47. Dacheux L, Moreau A, Ataman-Onal Y, Biron F, Verrier B, Barin F. 2004. Evolutionary dynamics of the glycan shield of the human immunodeficiency virus envelope during natural infection and implications for exposure of the 2G12 epitope. *J. Virol.* 78:12625–37

48. Kwong PD, Doyle ML, Casper DJ, Cicala C, Leavitt SA, et al. 2002. HIV-1 evades antibody-mediated neutralization through conformational masking of receptor-binding sites. *Nature* 420:678–82

49. Myszka DG, Sweet RW, Hensley P, Brigham-Burke M, Kwong PD, et al. 2000. Energetics of the HIV gp120-CD4 binding reaction. *Proc. Natl. Acad. Sci. USA* 97:9026–31

50. Wyatt R, Kwong PD, Desjardins E, Sweet RW, Robinson J, et al. 1998. The antigenic structure of the HIV gp120 envelope glycoprotein. *Nature* 393:705–11

51. Thali M, Furman C, Ho DD, Robinson J, Tilley S, et al. 1992. Discontinuous, conserved neutralization epitopes overlapping the CD4-binding region of human immunodeficiency virus type 1 gp120 envelope glycoprotein. *J. Virol.* 66:5635–41

52. Pantophlet R, Saphire EO, Poignard P, Parren PWHI, Wilson IA, Burton DR. 2003. Fine mapping of the interaction of neutralizing and nonneutralizing monoclonal antibodies with the CD4 binding site of human immunodeficiency virus type 1 gp120. *J. Virol.* 77:642–58

53. Center RJ, Lebowitz J, Leapman RD, Moss B. 2004. Promoting trimerization of soluble human immunodeficiency virus type 1 (HIV-1) Env through the use of HIV-1/simian immunodeficiency virus chimeras. *J. Virol.* 78:2265–76

54. Earl PL, Broder CC, Long D, Lee SA, Peterson J, et al. 1994. Native oligomeric human immunodeficiency virus type 1 envelope glycoprotein elicits diverse monoclonal antibody reactivities. *J. Virol.* 68:3015–26

55. Yang XZ, Farzan M, Wyatt R, Sodroski J. 2000. Characterization of stable, soluble trimmers containing complete ectodomains of human immunodeficiency virus type 1 envelope glycoproteins. *J. Virol.* 74:5716–25

56. Srivastava IK, Stamatatos L, Legg H, Kan E, Fong A, et al. 2002. Purification and characterization of oligomeric envelope glycoprotein from a primary R5 subtype B human immunodeficiency virus. *J. Virol.* 76:2835–47

57. VanCott TC, Veit SC, Kalyanaraman V, Earl P, Birx DL. 1995. Characterization of a soluble, oligomeric HIV-1 gp160 protein as a potential immunogen. *J. Immunol. Methods* 183:103–17

58. Zhang CW, Chishti Y, Hussey RE, Reinherz EL. 2001. Expression, purification, and characterization of recombinant HIV gp140. The gp41 ectodomain of HIV or simian immunodeficiency virus is sufficient to maintain the retroviral envelope glycoprotein as a trimer. *J. Biol. Chem.* 276:39577–85

59. Chakrabarti BK, Kong WP, Wu BY, Yang ZY, Friborg J, et al. 2002. Modifications of the human immunodeficiency virus envelope glycoprotein enhance immunogenicity for genetic immunization. *J. Virol.* 76:5357–68

60. Chen B, Zhou GF, Kim MY, Chishti Y, Hussey RE, et al. 2000. Expression, purification, and characterization of gp160e, the soluble, trimeric ectodomain of the simian immunodeficiency virus envelope glycoprotein, gp160. *J. Biol. Chem.* 275:34946–53

61. Farzan M, Choe H, Desjardins E, Sun Y, Kuhn J, et al. 1998. Stabilization of human immunodeficiency virus type 1 envelope glycoprotein trimers by disulfide bonds introduced into the gp41 glycoprotein ectodomain. *J. Virol.* 72:7620–25

62. Binley JM, Sanders RW, Clas B, Schuelke N, Master A, et al. 2000. A recombinant human immunodeficiency virus type 1 envelope glycoprotein complex stabilized by an intermolecular disulfide bond between the gp120 and gp41 subunits is an antigenic mimic of the trimeric virion-associated structure. *J. Virol.* 74:627–43

63. Sanders RW, Schiffner L, Master A, Kajumo F, Guo Y, et al. 2000. Variable-loop-deleted variants of the human immunodeficiency virus type 1 envelope glycoprotein can be stabilized by an intermolecular disulfide bond between the gp120 and gp41 subunits. *J. Virol.* 74:5091–100

64. Schulke N, Vesanen MS, Sanders RW, Zhu P, Lu M, et al. 2002. Oligomeric and conformational properties of a proteolytically mature, disulfide-stabilized human immunodeficiency virus type 1 gp140 envelope glycoprotein. *J. Virol.* 76:7760–76

65. Jeffs SA, Goriup S, Kebble B, Crane D, Bolgiano B, et al. 2004. Expression and characterization of recombinant oligomeric envelope glycoproteins derived from primary isolates of HIV-1. *Vaccine* 22:1032–46

66. Staropoli I, Chanel C, Girard M, Altmeyer R. 2000. Processing, stability, and receptor binding properties of oligomeric envelope glycoprotein from a primary HIV-1 isolate. *J.Biol. Chem.* 275:35137–45

67. Yang XZ, Florin L, Farzan M, Kolchinsky P, Kwong PD, et al. 2000. Modifications that stabilize human immunodeficiency virus envelope glycoprotein trimers in solution. *J.Virol.* 74:4746–54

68. Yang XZ, Lee J, Mahony EM, Kwong PD, Wyatt R, Sodroski J. 2002. Highly stable trimers formed by human immunodeficiency virus type 1 envelope glycoproteins fused with the trimeric motif of T4 bacteriophage fibritin. *J. Virol.* 76:4634–42

69. Chen B, Cheng YF, Calder L, Harrison SC, Reinherz EL, et al. 2004. A chimeric protein of simian immunodeficiency virus envelope glycoprotein gp140 and *Escherichia coli* aspartate transcarbamoylase. *J. Virol.* 78:4508–16

70. Yang XZ, Wyatt R, Sodroski J. 2001. Improved elicitation of neutralizing antibodies against primary human immunodeficiency viruses by soluble stabilized envelope glycoprotein trimers. *J. Virol.* 75:1165–71

71. Grundner C, Li YX, Louder M, Mascola J, Yang XZ, et al. 2005. Analysis of the neutralizing antibody response elicited in rabbits by repeated inoculation with trimeric HIV-1 envelope glycoproteins. *Virology* 331:33–46

72. Earl PL, Sugiura W, Montefiori DC, Broder CC, Lee SA, et al. 2001. Immunogenicity and protective efficacy of oligomeric human immunodeficiency virus type 1 gp140. *J.Virol.* 75:645–53

73. Kim M, Qiao ZS, Montefiori DC, Haynes BF, Reinherz EL, Liao HX. 2005. Comparison of HIV type 1 ADA gp120 monomers versus gp140 trimers as immunogens for the induction of neutralizing antibodies. *AIDS Res. Hum. Retrovir.* 21:58–67

74. Rossio JL, Esser MT, Suryanarayana K, Schneider DK, Bess JW Jr, et al. 1998. Inactivation of human immunodeficiency virus type 1 infectivity with preservation of conformational and functional integrity of virion surface proteins. *J. Virol.* 72:7992–8001

75. Arthur LO, Bess JW Jr, Chertova EN, Rossio JL, Esser MT, et al. 1998. Chemical inactivation of retroviral infectivity by targeting nucleocapsid protein zinc fingers: a candidate SIV vaccine. *AIDS Res. Hum. Retrovir.* 14(Suppl. 3):S311–19

76. Grundner C, Mirzabekov T, Sodroski J, Wyatt R. 2002. Solid-phase proteoliposomes containing human immunodeficiency virus envelope glycoproteins. *J. Virol.* 76:3511–21

77. Lifson JD, Rossio JL, Piatak M Jr, Bess J Jr, Chertova E, et al. 2004. Evaluation of the safety, immunogenicity, and protective efficacy of whole inactivated simian immunodeficiency virus (SIV) vaccines with conformationally and functionally intact envelope glycoproteins. *AIDS Res. Hum. Retrovir.* 20:772–87

78. Poon B, Safrit JT, McClure H, Kitchen C, Hsu JF, et al. 2005. Induction of humoral immune responses following vaccination with envelope-containing, formaldehyde-treated, thermally inactivated human immunodeficiency virus type 1. *J. Virol.* 79:4927–35

79. Doan LX, Li M, Chen CY, Yao QZ. 2005. Virus-like particles as HIV-1 vaccines. *Rev. Med. Virol.* 15:75–88

80. Montefiori DC, Safrit JT, Lydy SL, Barry AP, Bilska M, et al. 2001. Induction of neutralizing antibodies and gag-specific cellular immune responses to an R5 primary isolate of human immunodeficiency virus type 1 in rhesus macaques. *J. Virol.* 75:5879–90

81. Dale CJ, Liu XS, De Rose R, Purcell DF, Anderson J, et al. 2002. Chimeric human papilloma virus-simian/human immunodeficiency virus virus-like-particle vaccines: immunogenicity and protective efficacy in macaques. *Virology* 301:176–87

82. Chakrabarti BK, Kong WP, Wu B-Y, Yang Z-Y, Friborg J Jr, Ling X, King SR, Montefiori DC, Nabel GJ. 2002. Modifications of the human immunodeficiency virus envelope glycoprotein enhance immunogenicity for genetic immunization. J Virol 76: 5357–5368.

83. Pancera M, Majeed S, Ban YE, Chen L, Huang CC, Kong L, Kwon YD, Stuckey J, Zhou T, Robinson JE, et al. 2010a. Structure of HIV-1 gp120 with gp41-interactive region reveals layered envelope architecture and basis of conformational mobility. Proc Natl Acad Sci 107: 1166–1171.

84. Wu X, Yang ZY, Li Y, Hogerkorp CM, Schief WR, Seaman MS, Zhou T, Schmidt SD, Wu L, Xu L, et al. 2010. Rational design of envelope identifies broadly neutralizing human monoclonal antibodies to HIV-1. Science 329: 856–861.

85. Agrawal-Gamse C, Luallen RJ, Liu B, Fu H, Lee FH, Geng Y, Doms RW. 2011. Yeast elicited cross-reactive antibodies to HIV Env glycans efficiently neutralize virions expressing exclusively high-mannose N-linked glycans. J Virol 85: 470–480

86. Walker LM, SimekMD, Priddy F, Gach JS,Wagner D, Zwick MB, Phogat SK, Poignard P, Burton DR. 2010b. A limited number of antibody specificities mediate broad and potent serum neutralization in selected HIV-1 infected individuals. PLoS Pathog 6: e1001028

87. Ofek G,Guenaga FJ, Schief WR, Skinner J, BakerD,Wyatt R, Kwong PD. 2010a. Elicitation of structure-specific antibodies by epitope scaffolds. Proc Natl Acad Sci 107:17880–17887.

88. Dey B, Svehla K, Xu L,Wycuff D, Zhou T, Voss G, Phogat A, Chakrabarti BK, Li Y, ShawG, et al. 2009. Structure-based stabilization of HIV-1 gp120 enhances humoral immune responses to the induced co-receptor binding site. PLoS Pathog 5: e1000445.

89. Flower R.D. Bioinformatics for vaccinology, 2008 by John Wiley & Sons, Ltd.

90. Greenbaum J, Andersen P, Blythe M, Bui H, Cachau R, Crowe J, Davies M, Kolaskar A, Lund O, Morrison S, et al: Towards a consensus on datasets and evaluation metrics for developing B-cell epitope prediction tools. J.Mol. Recognit 2007, 20:75-82.

91. Yasser EL-Manzalawy1,2*, Vasant Honavar2: Recent advances in B-cell epitope prediction Methods. Immunome Research 2010, 6(Suppl 2):S2

92. Parker J, Guo HRD: New hydrophilicity scale derived from highperformance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and x-ray-derived accessible sites. Biochemistry 1986, 25:5425 5432.

93. Karplus P, Schulz G: Prediction of chain flexibility in proteins: a tool for the selection of peptide antigen. Naturwiss 1985, 72:21-213.

94. Pellequer J, Westhof E, Van Regenmortel M: Correlation between the location of antigenic sites and the prediction of turns in proteins. Immunol. Lett. 1993, 36:83-99.

95. Emini E, Hughes J, Perlow D, Boger J: Induction of hepatitis A virus neutralizing antibody by a virus-specific synthetic peptide. J. Virol. 1985, 55:836-839.

96. Hopp T, Woods K: Prediction of protein antigenic determinants from amino acid sequences. Proceedings of the National Academy of Sciences of the United States of America 1981, 78(6):3824.

97. Levitt M: A simplified representation of protein conformations for rapid simulation of protein folding. Journal of molecular biology 1976, 104:59.

98. Pellequer J, Westhof E, Van Regenmortel M: Predicting location of continuous epitopes in proteins from their primary structures. Meth. Enzymol. 1991, 203:176-201.

99. Alix A: Predictive estimation of protein linear epitopes by using the program PEOPLE. Vaccine 1999, 18:311-4.

100. Odorico M, Pellequer J: BEPITOPE: predicting the location of continuous epitopes and patterns in proteins. J. Mol. Recognit. 2003, 16:20-22.

101. Saha S, Raghava G: BcePred: Prediction of continuous B-cell epitopes in antigenic sequences using physico-chemical properties. Artificial Immune Systems, Third International Conference (ICARIS 2004), LNCS 2004, 3239:197-204.

102. EL-Manzalawy Y, Dobbs D, Honavar V: Predicting linear B-cell epitopes using string kernels. J. Mol. Recognit. 2008, 21:243-255.

103. EL-Manzalawy Y, Dobbs D, Honavar V: Predicting flexible length linear Bcell epitopes. 7th International Conference on Computational Systems Bioinformatics 2008, 121-131.

104. Saha S, Raghava G: Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. Proteins 2006, 65:40-48.

105. Chen C, Zhou X, Tian Y, Zou X, Cai P: Predicting protein structural class with pseudo-amino acid composition and support vector machine fusion network. Analytical biochemistry 2006, 357:116-121.

106. Wu J, Liu H, Duan X, Ding Y, Wu H, Bai Y, Sun X: Prediction of DNAbinding residues in proteins from amino acid sequences using a random forest model with a hybrid feature. Bioinformatics 2009, 25:30.

107. Yan C, Terribilini M, Wu F, Jernigan R, Dobbs D, Honavar V: Predicting DNA binding sites of proteins from amino acid sequence. BMC bioinformatics 2006, 7:262.

108. Terribilini M, Lee J, Yan C, Jernigan R, Honavar V, Dobbs D: Prediction of RNA binding sites in proteins from amino acid sequence. Rna 2006, 12(8):1450-1462.

109. Kumar M, Gromiha M, Raghava G: Prediction of RNA binding sites in a protein using SVM and PSSM profile. Proteins 2008, 71:189-194.

110. Kulkarni-Kale U, Bhosle S, Kolaskar A: CEP: a conformational epitope prediction server. Nucleic Acids Res. 2005, 33:W168.

111. Haste Andersen P, Nielsen M, Lund O: Prediction of residues in discontinuous B-cell epitopes using protein 3D structures. Protein Sci.2006, 15:2558.

112. Mitchell J, Kerr R, Ten Eyck L: Rapid atomic density methods for molecular shape characterization. J Mol Graph Model 2001, 19(3-4):325-330.

113. Camacho C, Zhang C: FastContact: rapid estimate of contact and binding free energies. Bioinformatics 2005, 21(10):2534-2536

114.   Moreau V, Granier C, Villard S, Laune D, Molina F: Discontinuous epitope prediction based on mimotope analysis. Bioinformatics 2006, 22(9):1088-1095.

115.   Bublil E, Freund N, Mayrose I, Penn O, Roitburd-Berman A, Rubinstein N, Pupko T, Gershoni J: Stepwise prediction of conformational discontinuous B-cell epitopes using the Mapitope algorithm. Proteins 2007, 68:294-304.

116.   Mayrose I, Penn O, Erez E, Rubinstein N, Shlomi T, Freund N, Bublil E, Ruppin E, Sharan R, Gershoni J, et al: Pepitope: epitope mapping from affinity-selected peptides. Bioinformatics 2007, 23(23):3244-3246.

117.   Batori, V. et al. (2006) An in silico method using an epitope motif database for predicting the location of antigenic determinants on proteins in a structural context. J. Mol. Recogn. 19, 21–29

118.   Huang, Y.X. et al. (2008) Pep-3D-Search: a method for B-cell epitope prediction based on mimotope analysis. BMC Bioinform. 9, 538 10.1186/1471-2105-9-538

119.   Mihai L. Azoitei 1, Yih-En Andrew Ban 1, Jean-Philippe Julien 2, Steve Bryson 2, 3, Alexandria Schroeter 1, Oleksandr Kalyuzhniy 1, Justin R. Porter 1, 4, Yumiko Adachi 1, David Baker 1, Emil F. Pai 2, 3, 5 and William R. Schief 1, 6, 7. Computational Design of High-Affinity Epitope Scaffolds by Backbone Grafting of a Linear Epitope. J. Mol. Biol. (2012) 415, 175–192

120.   Mihai L. Azoitei, Bruno E. Correia, Yih-En Andrew Ban, Chris Carrico, Oleksandr Kalyuzhniy, Lei Chen, Alexandria Schroeter, Po-Ssu Huang, Jason S. McLellan, Peter D. Kwong, David Baker, Roland K. Strong, William R. Schief*, Computation-Guided Backbone Grafting of a Discontinuous Motif onto a Protein Scaffold. Published 21 October 2011, *Science* 334, 373 (2011)

121.   Bruno E. Correia,[1,2,6] Yih-En Andrew Ban,[1,6,7] Margaret A. Holmes,[3] Hengyu Xu,[3] Katharine Ellingson,[4] Zane Kraft,[4] Chris Carrico,[1] Erica Boni,[3] D. Noah Sather,[4] Camille Zenobia,[3] Katherine Y. Burke,[3] Tyler Bradley-Hewitt,[3] Jessica F. Bruhn-Johannsen,[3] Oleksandr Kalyuzhniy,[1] David Baker,[1] Roland K. Strong,[3] Leonidas Stamatatos,[4,5] and William R. Schief, Computational Design of Epitope-Scaffolds Allows Induction of Antibodies Specific for a Poorly Immunogenic HIV Vaccine Epitope. j.str.2010.06.010

122.   Wayne C. Koff, HIV vaccine development: Challenges and opportunities towards a solution of the HIV vaccine-neutralizing antibody problem. JVAC 12560 1–6 Elsevier

123.   Andrew J McMichael & Barton F Haynes, Lessons learned from HIV-1 vaccine trials: new priorities and directions. nature immunology volume 13 number 5 may 2012.

124.   29. Gray, E.S. *et al. J. Virol.* **85**, 4828–4840 (2011).

125.   Diskin, R. *et al. Science* **334**, 1289–1293 (2011).

126.   Mouquet, H. *et al. Nature* **467**, 591–595 (2010).

127.   Walker, L.M. *et al. Nature* **477**, 466–470 (2011).

128.   Julie Overbaugh and Lynn Morris, The Antibody Response against HIV-1. *Cold Spring Harb Perspect Med* 2012; doi: 10.1101/cshperspect.a007039

129.   Dreja H, O'Sullivan E, Pade C, Greene KM, Gao H, Aubin K, Hand J, Isaksen A, D'Souza C, Leber W, et al. 2010. Neutralization activity in a geographically diverse East London cohort of human immunodeficiency virus type 1-infected patients: Clade C infection results in a stronger and broader humoral immune response than clade B infection. J Gen Virol 91: 2794–2803.

130.   Haynes F.B. et al, B-cell–lineage immunogen design in vaccine development with HIV-1 as a case study. Nature biotechnology volume 30 number 5 may 2012

131.   Kwong et al, Structural Basis for Broad and Potent Neutralization of HIV-1 by Antibody VRC01. Science. 2010 August 13

132.   Chen L, et al. Science. 2009; 326:1123
133.   Maddon PJ, et al. Cell. Aug.1985 42:93.
134.   Kwong PD, et al. Nature. Jun 18.1998 393:648.
135.   Pancera M, et al. Proc Natl Acad Sci U S A. Jan 19.2010 107:1166.
136.   Rits-Volloch S, Frey G, Harrison SC, Chen B. EMBO J. Oct 18.2006 25:5026.
137.   Wu X, et al. J Virol. Nov.2009 83:10892.
138.   Zhou T, et al. Nature. Feb 15.2007 445:732
139.   Wrin T, Nunberg JH. 1994. HIV-1MN recombinant gp120 vaccine serum, which fails to neutralize primary isolates of HIV-1, does not antagonize neutralization by antibodies from infected individuals. AIDS 8: 1622–1623.
140.   Flynn NM, Forthal DN, Harro CD, Judson FN, Mayer KH, Para MF. 2005. Placebo-controlled phase 3 trial of a recombinant glycoprotein 120 vaccine to prevent HIV-1 infection. J Infect Dis 191: 654–665

141.   Gilbert PB, Peterson ML, Follmann D, Hudgens MG, Francis DP, Gurwith M, Heyward WL, Jobes DV, Popovic V, Self SG, et al. 2005. Correlation between immunologic responses to a recombinant glycoprotein 120 vaccine and incidence of HIV-1 infection in a phase 3 HIV-1 preventive vaccine trial. J Infect Dis 191: 666–677.

142.   Rerks-Ngarm S, Pitisuttithum P, Nitayaphan S, Kaewkungwal J, Chiu J, Paris R, Premsri N, Namwat C, de Souza M, Adams E, et al. 2009. Vaccination with ALVAC and AIDSVAX to prevent HIV-1 infection in Thailand. N Engl J Med 361: 2209–2220.

143.   Forthal, D.N., Gilbert, P.B., Landucci, G. & Phan, T. *J. Immunol.* **178**, 6596–6603 (2007).

144.   Walker LM, et al. 2009. Broad and potent neutralizing antibodies from an African donor reveal a new HIV-1 vaccine target. Science 326: 285–289

145.   Trkola A, Purtscher M, Muster T, Ballaun C, Buchacher A, Sullivan N, Srinivasan K, Sodroski J, Moore JP, Katinger H. 1996. Human monoclonal antibody 2G12 defines a distinctive neutralization epitope on the gp120 glycoprotein of human immunodeficiency virus type 1. J Virol 70: 1100–1108.

146.   Walker LM, Chan-Hui P. 2010. High through-put functional screening of activated B cells from 4 African elite neutralizers yields a panel of novel broadly neutralizing antibodies. In Abstracts from AIDS Vaccine 2010, pp. A-149–A-150. Atlanta, Georgia.

147.   Corti D, Langedijk JP, Hinz A, Seaman MS, Vanzetta F, Fernandez-Rodriguez BM, Silacci C, Pinna D, Jarrossay D, Balla-Jhagjhoorsingh S, et al. 2010. Analysis of memory B cell responses and isolation of novel monoclonal antibodies with neutralizing breadth from HIV-1-infected individuals. PLoS One 5: e8805.

148.   Sabin C, Corti D, Buzon V, Seaman MS, Lutje Hulsik D, Hinz A, Vanzetta F, Agatic G, Silacci C, Mainetti L, et al. 2010. Crystal structure and size-dependent neutralization properties of HK20, a human monoclonal antibody binding to the highly conserved heptad repeat 1 of gp41. PLoS Pathog 6: e1001195.

149.   Rose et al.The RCSB Protein Data Bank: redesignedweb site and web services. Nucleic Acids Research, 2011, Vol. 39, Database issue