Title: A facile method of mapping neutralizing HIV-1 epitopes

Authors: Rohini Datta[1,§], Rohan Roy Chowdhury[1], Kavyashree Manjunath[2], Luke Elizabeth Hanna[3] and Raghavan Varadarajan[1,4,*]

Affiliations: [1]Molecular Biophysics Unit, Indian Institute of Science, Bangalore-560012, India; [2]Institute for Stem Cell Science and Regenerative Medicine, Bangalore-560065, India; [3]Department of HIV/AIDS, National Institute for Research in Tuberculosis, Indian Council of Medical Research, Chennai-600031, India; [4]Jawaharlal Nehru Centre for Advanced Scientific Research, Bangalore-560004, India; [§]Present Address: Department of Genetics, Stanford School of Medicine, Stanford University, USA

*__Corresponding Author:__ Raghavan Varadarajan, E-mail: varadar@iisc.ac.in

Abstract

Identification of specific epitopes targeted by neutralizing antibodies is essential to advance epitope-based vaccine design strategies. We report a methodology for rapid epitope mapping of neutralizing antibodies against HIV-1 Env at single residue resolution, using viral neutralization assays and deep sequencing. We extended our methodology to map multiple specificities of epitopes targeted in polyclonal sera, elicited in immunized animals as well as in HIV-1 infected patients.

Main

We recently described a methodology to decipher epitopes of monoclonal antibody panels using yeast surface display, Cys labeling and deep sequencing [1, 2]. Here we built on this to map epitopes of neutralizing monoclonal antibodies and polyclonal sera against the HIV-1 virus. The design of an effective vaccine against HIV-1 has met with limited success so far, because of the inability of immunogens to elicit broadly neutralizing antibodies (bNAbs) targeted to the trimeric envelope (Env) glycoprotein, the sole viral antigen on the surface of the virus. Most bNAbs isolated from infected individuals are directed to one of the major sites of vulnerability on Env: V1/V2 loop apex, V3 loop, CD4-binding site, centre of the gp120 silent face, gp120-gp41 subunit interface, gp41 fusion peptide and membrane proximal external region (MPER) of gp41.

We developed a facile methodology for mapping neutralizing HIV-1 epitopes at single residue resolution. To this end, a library of single-site Cys mutations were engineered at selected solvent-exposed sites in the Env glycoprotein on the surface of the virus. Cys mutants were covalently labeled using a Cys reactive reagent (Maleimide-PEG_2-Biotin) that masks epitope residues, followed by infection of the labeled mutant viruses in mammalian cells in the presence of bNAbs. Subsequently, deep sequencing of the env gene from bNAb-resistant viruses was used to delineate epitopes (Figure 1). We focused our epitope mapping efforts to the Env ectodomain of the clade B, CCR5-tropic primary HIV-1 isolate JRFL, which has been extensively characterized both biochemically and structurally, and is used as a model primary virus.

A total of 81 solvent-exposed residues (Table S1) were selected from the X-ray crystal structure of the pre-fusion HIV-1 Env ectodomain using a combined criterion of >30% solvent accessibility and sidechain-sidechain centroid distance >8 Å (Figure S1). This led to the

identification of a set of solvent-exposed residues which collectively exhibit uniform surface coverage of the Env trimer (Figure S2). These Env residues were mutated to Cys in the pLAI-JRFL molecular clone, pooled in equimolar quantities to create a plasmid DNA library, and transfected in HEK293T cells to produce the Exposed Cys Library, a library of single-site Cys mutant viruses. Wt Env lacks free Cys residues. In single-cycle and multi-cycle infectivity assays, the Exposed Cys Library retained significant viral infectivity (Figure S3) and was therefore used for epitope mapping studies.

Viral supernatants from HEK293T cells contain 20% Fetal Bovine Serum (FBS) which has a high protein content (~30-40 mg/ml) and includes a large number of sulfhydryl groups that can be labeled. This potentially necessitated the use of high concentrations of the Cys label for our experiments. We exploited the high affinity interaction of the bacterial toxin CcdB and DNA Gyrase[3] to quantify the extent of Cys labeling in the presence of 20% FBS, to mimic viral labeling (Figure S4A). The use of 10 mM Cys label and overnight incubation led to the maximum extent of labeling (Figure S4B). Next, we covalently labeled the WT virus with Maleimide-PEG_2-Biotin, and measured single-cycle viral infectivity in TZM-bl cells. Since the use of 10 mM label likely led to a high extent of label incorporation without adversely affecting viral infectivity (Figure S5), we used this concentration to label the WT virus and the Exposed Cys Library.

To validate our methodology, proof-of-concept studies were carried out with the Exposed Cys Library to map epitopes of bNAbs directed to distinct sites on Env. To decide the amount of antibody to use, neutralization assays of WT LAI-JRFL virus were performed in TZM-bl cells, in the presence of: CD4-binding site bNAb VRC01, V3-glycan targeting bNAb PGT128 and gp120-gp41 interface specific bNAb PGT151. For epitope mapping studies, we used antibody concentrations 100 times in excess of the calculated IC_{50} (Table S2).

Mapping epitopes using the Exposed Cys Library necessitated env gene sequencing from bNAb-resistant virion pools (Figure 1). Isolation of env clones was accomplished using HUT-R5, a T-cell line permissive to multi-cycle infection. To map known bNAb epitopes using our methodology, we carried out multi-cycle infectivity assays of the labeled Exposed Cys Library in the presence of 100 IC_{50} equivalents (and subsequent threefold dilutions) of bNAbs VRC01, PGT128 and PGT151. Multi-cycle infectivity of the labeled Exposed Cys Library was found to be similar in the absence and presence of the bNAb VRC01 (Figure S6).

Since epitopes for the selected bNAbs are present in the Exposed Cys Library, some mutant viruses are expected to be resistant to neutralization upon bNAb exposure. This was inferred by quantifying the ratio of extracellular p24 levels for the labeled Exposed Cys Library in the absence to presence of VRC01 and other bNAbs. The ratio of p24 levels was found to be ~1 for the labeled virus in the absence to presence of bNAbs (Figure 2). This indicated that mutant viruses became resistant to neutralization and proved that epitopes for these bNAbs are present in the Exposed Cys Library, and can be mapped.

Delineating anti-HIV polyclonal antibody specificities using standard epitope mapping techniques is challenging since the polyclonal antibody repertoire is most often comprised of bNAbs that target multiple epitopes on Env. We therefore extended our methodology to characterize epitope specificities of polyclonal sera elicited by HIV-1 infected individuals and immunized animals. Plasma from a subtype-C HIV-infected elite neutralizer (NAB059) was used for epitope mapping studies using our approach[4, 5]. Epitope mapping was also carried out using sera from guinea pigs immunized with trimeric cyclic permutants of gp120[6] which were previously shown to exhibit Tier 2 neutralization against a global panel of HIV-1 isolates[7].

Multi-cycle infectivity assays of the labeled Exposed Cys Library were carried out in HUT-R5 cells, with 100 ID_{50} equivalents of NAB059 plasma or 10 ID_{50} equivalents of trimeric cyclic

permutant-immunized guinea pig sera. Both the NAB059 plasma and guinea pig sera were found to select for neutralization resistant viruses in the labeled Exposed Cys Library (Figure S7), indicating that epitopes for these polyclonal antibodies are present in the library.

Deep sequencing of the env gene recovered from labeled virion pools in the presence and absence of antibody/plasma/sera was subsequently carried out to map epitopes at single residue resolution. The env gene was sequenced in 6 fragments with PCR primers containing unique MID tags representing a particular selection condition: bNAb, plasma/sera or controls (Table S3). A total of 69 Cys mutants were identified from the deep sequencing data. Of the 12 missing mutants, 10 mutants (D78C, R151C, N160C, N187C, E267C, E293C, D340C, N392C, N398C, R440C) were not present in the plasmid DNA pool (no reads for MID 8) used for transfection, and were likely lost while pooling. A couple of mutants (N88C and Q442C) were present in the plasmid DNA pool (MID 8), but showed complete loss of infectivity upon Cys labeling (no reads for MID 6). The number of reads for each Cys mutant in the presence of the antibody/plasma/sera was normalized with respect to the corresponding number in the absence as described in the Methods section. For virtually all mutants outside the epitope(s), this normalized read ratio (R) is close to zero because antibody mediated neutralization eliminates these mutants. Residues with $R > 0.5$ were designated as epitope residues. Epitopes of bNAbs identified using our methodology were in excellent agreement with available structural information (Figure 3, Table S4). Polyclonal NAB059 plasma contained quaternary epitope-specific and CD4-binding site-specific signatures (Table S4). The cyclic permutant immunized sera contained epitope specificities mapping to the V1V2 loop and CD4-binding site (Table S4), consistent with earlier observations using mammalian cell surface display and mutant pseudoviruses[8, 9]. Both polyclonal sera were observed to encompass multiple antibody specificities, which account for their observed neutralization activity.

Finally, we validated our epitope mapping strategy on a panel of epitope mutants of the bNAbs from our deep sequencing data (Table S5). Individual viruses containing a Cys mutation either at or outside a bNAb epitope in Env were labeled with the Cys label, and viral infectivity assays were performed in TZM-bl cells in the presence and absence of bNAbs. Titration of the viral supernatant was carried out using both qRT-PCR (Figure S8, Table S6) as well as TZM-bl cell-based reporter assays (Figure S9).

Viral infectivity assays of Cys mutant viruses were carried out in the presence of 100 IC_{50} equivalents (and threefold dilutions) of bNAbs VRC01, b12, PGT128 and PGT151. Viruses became resistant to neutralization only when the corresponding epitopes were masked by the bulky Cys label (Figure S10). Control experiments were carried out by labeling the epitope-Cys mutant virus in the presence of bNAbs targeting other sites on Env (Figure S11). The results clearly demonstrate that the methodology was able to distinguish between epitope and non-epitope residues for each bNAb.

High-throughput, site-directed scanning mutagenesis techniques are widely used in protein engineering, which enable studies of protein structure and function at single residue resolution[1, 2, 10, 11]. A deep mutational scanning-based approach, "mutational antigenic profiling" was recently used to map epitopes of nine HIV-1 bNAbs[12]. This approach is based on the creation of codon-mutant libraries[13], identification of bNAb escape mutants and deep sequencing of virus pools to quantify the enrichment of mutations after antibody selection. This method comprehensively delineates regions of viral escape or functional epitopes, not all of which are in physical contact with the antibody[12].

We use site-directed Cys scanning mutagenesis, a powerful tool to probe protein function [14-17]. Cys is an attractive candidate for mutagenesis since Cys at solvent-exposed sites in proteins can be labeled site-specifically. In addition, Cys is relatively infrequent in the proteome, and

introduction of Cys in proteins is most often accompanied by very little structural and functional perturbation[18-20]. Moreover, our method is devoid of the complexity of saturation mutagenesis libraries, which makes library construction and analysis of deep sequencing data considerably simpler. Our library sizes are at least 20-60 times smaller than saturation mutagenesis libraries. Our method appears to exclusively identify structurally defined epitope residues that are in close physical proximity to the antibody (Figure 3). In contrast to recently developed cryo-EM based epitope-mapping approaches[21], the method does not require sophisticated instrumentation, is highly parallelizable and can map epitopes to single residue resolution.

These features make it an attractive approach to map neutralization responses in human sera from either HIV-1 infected individuals or vaccinees. Knowledge of the three-dimensional structure of the antigen is not a prerequisite for selecting residues for mutagenesis using the proposed methodology. In the absence of structural information, homology modeling or Asp scanning mutagenesis [22] can be employed to locate solvent accessible residues in proteins.

There is an urgent need for reliable epitope mapping techniques to elucidate neutralizing epitopes against HIV-1 and others viruses, including SARS-CoV-2. Deep sequencing of the env gene recovered from labeled virion pools in the presence of a bNAbs/sera, using our methodology provides comprehensive knowledge of epitope specificities, which in turn provides useful insights about the nature of antibody response elicited by the virus[23-25] and informs immunogen design[26, 27]. The methodology can readily be coupled to cell surface display methodologies[2] to obtain all, rather than just neutralization epitopes.

Methods.

Plasmids.

Site-directed Cys scanning mutagenesis was carried out in the env gene (~2.2 kb) of HIV-1 molecular clone pLAI-JRFL (11.7 kb), where LAI-JRFL refers to the strain of HIV-1. Since site-directed mutagenesis is less efficient for larger plasmids (> 3.1 kb), the env gene was subcloned in a smaller TA cloning vector pTZ57/R using the insTA clone PCR cloning kit (Thermo Fisher Scientific).

bNAbs.

All bNAbs were obtained either from the NIH AIDS reagent program (https://www.aidsreagent.org/) or the IAVI Neutralizing Antibody Consortium.

Residue selection for Exposed Cys Library.

A combined criterion of >30% solvent accessibility and sidechain-sidechain centroid distance >8 Å was used to select 81 residues from the X-ray crystal structure of the pre-fusion HIV-1 Env ectodomain (PDB ID: 4TVP). The solvent accessibility of each residue was calculated using the NACCESS V 2.1.1 program (http://wolf.bms.umist.ac.uk/naccess/) after removing antibody chains. Pairwise centroid distances between side chain centroids were calculated for all residues exhibiting solvent accessibility >30% using an in-house PERL script. This sidechain-sidechain centroid distance cut-off was chosen to leave out residues in contact[3]. The 81 selected residues occupy a surface area roughly corresponding to one-third of the total surface area of the Env ectodomain (PDBePISA, EMBL-EBI). In addition, 47 out of the 83 selected exposed residues are contact residues either for CD4, the primary cellular receptor for HIV, or the following bNAbs directed to various neutralizing epitopes on the Env: CD4-binding site (VRC01, PGV04, 3BNC117, NIH45-46, CH103, b12), CD4-induced epitope (17b), V3 glycan (PGT121-123, 125-128, 130, 131, 135), V1V2 glycan (PG9, CAP256-

VRC26), MPER (2F5), gp120-gp41 interface (PGT151) and mannose-dependent epitope (2G12).

Cys library generation.

To generate the Cys library, 81 selected exposed residues listed in Table S1 were individually mutated to Cys by site-directed mutagenesis using partially overlapping complementary primers which generate a doubly-nicked amplicon[28]. There was a 21 bp overlap between the forward and reverse partially overlapping primers, and a Cys codon was located 9 bp from the 5'-end of each primer. All primers were synthesized at the Protein and Nucleic Acid (PAN) facility at Stanford University, USA.

Site-directed mutagenesis reactions were carried out using Phusion DNA polymerase (NEB) according to the manufacturer's protocol in a thermal cycler PTC 200 (Bio-Rad). Each 20 µl reaction had 20-25 ng of template (pTZ57/R-env), 1 µM of the primer pair, 1X GC Buffer, 200 µM dNTPs and 3% DMSO as an additive. PCR was done for 30 cycles with an annealing temperature of 65°C for 35 seconds. The reaction mixture was subjected to DpnI digestion by adding 0.5 µl DpnI (20 U/µl), 1 µl NEB Buffer 4 and 8.5 µl of the PCR product, followed by incubation at 37°C for 6 hours. PCR products were purified using the GeneJET PCR Purification kit (Thermo Fisher Scientific). Final elution of DNA was done in 20 µl MilliQ water. An aliquot of 1 µl of the purified PCR product was used to transform *E.coli* TOP10 electrocompetent cells and following recovery, inoculated on Luria-Bertani plates containing 100 µg/ml Ampicillin. Plasmid was isolated from using commercially available mini-prep columns (Qiagen). Incorporation of the Cys mutation at the selected positions in env was confirmed by Sanger sequencing, carried out at Macrogen, South Korea.

Single-Cys mutant pTZ57/R-env plasmids were pooled in equimolar quantities to create a pooled plasmid DNA library. The env gene harbouring Cys mutations was amplified from

pTZ57/R-env plasmid pools using Phusion DNA Polymerase (NEB) and primers from the 5' and 3' ends of the gene to amplify the entire env, and subsequently cloned into the pLAI-JRFL molecular clone between BamHI and SfiI restriction sites using T4 DNA ligase (NEB). This was accomplished by replacement of the WT env gene from the pLAI-JRFL molecular clone by digestion with BamHI and SfiI restriction enzymes, and subsequent purification of the vector backbone fragment by electroelution at 100 volts in 0.5X TAE for 1 hour 30 minutes. Following electroelution, the DNA fragment was extracted using phenol-chloroform-isoamyl alcohol (25:24:1), precipitated using Pellet Paint Co-Precipitant (EMD Millipore), and subsequently, ligation and electrotransformation in *E.coli* TOP10 were carried out. The mutant plasmid DNA library was purified using commercially available mini-prep columns (Qiagen) using growth conditions of 30°C and 130 rpm to support the propagation of large plasmids.

Transient production of Cys mutant viruses.

The Exposed Cys viral Library was produced by transient transfection of the pLAI-JRFL mutant plasmid DNA library into HEK293T cells. 24 hours before transfection, $5x10^5$ cells were seeded in each well of a 6-well plate, containing 2 ml of DMEM with 10% Fetal Bovine Serum (FBS). The transfection mix contained 16 µg of transfection reagent, polyethyleneimine (PEI) (Sigma-Aldrich) and 4 µg of mutant pLAI-JRFL plasmid (4:1 ratio of PEI:DNA), in a final volume of 100 µl of serum-free DMEM. After 10 minutes of incubation at room temperature, 100 µl of 10% DMEM was added to the transfection mix, and this was used to transfect each well of a 6-well plate containing HEK293T cells at 50-70% confluency. Media was replaced 6-8 hours post-transfection with 1.5-1.8 ml of 10% DMEM, and tissue culture media was collected 48 hours post-transfection to harvest the virus. Culture media was centrifuged at 1500 rpm for 10 minutes to remove residual cells. To the supernatant containing viruses, FBS was added to a final concentration of 20%. Viral supernatants were aliquoted and stored at -70°C. Titration of the viral supernatant was carried out using TZM-bl cell-based

reporter assays to determine viral infectivity. All studies with replication competent viruses were carried out at the Biosafety Level-3 facility at the Centre for Infectious Disease Research (CIDR) at IISc.

Single-cycle viral infectivity assay in TZM-bl cells.

Single-cycle infectivity assays of viral libraries harvested from HEK293T culture supernatants were carried out using a luciferase reporter gene-based assay in TZM-bl cells. TZM-bl (previously known as JC53BL-13) is a CXCR4-positive HeLa cell derivative, engineered to constitutively express the primary cellular receptor of HIV, CD4 and the CCR5 co-receptor. These cells are therefore permissive to infection by most HIV strains. Integrated into these cells is a firefly luciferase reporter gene under the control of an HIV-1 LTR promoter. Upon viral infection, the viral Tat protein activates the LTR promoter, and consequently, the extent of infection can be directly quantified by luminescence measurements from TZM-bl cells. For the assay, 100µl of undiluted and subsequent threefold-dilutions (in 10% DMEM) of viral supernatant was added in successive wells of a 96-well tissue culture plate. To this, $1x10^4$ freshly-trypsinized TZM-bl cells in 100 µl of 10% DMEM containing 30 µg/ml of DEAE-dextran was added and incubated for 48 hours. Subsequently, a 100 µl volume was discarded from each well and 80µl of reconstituted Britelite Plus reagent (PerkinElmer) was added and incubated for 2 minutes at room temperature. The reagent carries out cell lysis as well as the luciferase enzymatic reaction. Following incubation, 130 µl was aspirated from each well and transferred to a 96-well black plate and luminescence readings were measured using a VICTOR X2 luminometer (PerkinElmer).

Quantification of the extent of Cys labeling using MALDI-TOF.

WT CcdB and the test protein, GyrA14-I491C were purified using previously standardized protocols[29]. GyrA14-I491C was dissolved at 50 µM in 1X PBS buffer containing 20% FBS (to

mimic virus labeling) at pH 7.0-7.5, treated with 0.5mM TCEP to reduce disulfide bonds, and labeled with a commercially available Cys reactive reagent (CRR), EZ-Link™ Maleimide-PEG_2-Biotin (Thermo Fisher Scientific). At pH 7.0, the maleimide group is almost 1000 times more reactive towards free sulfhydryl groups as opposed to primary amines in the protein [30]. Labeling was carried out for different concentrations of the label, i.e. 0.01 mM, 0.1 mM, 1 mM and 10 mM and different time-points of incubation i.e., 15 mins, 1 hour, 4 hours and overnight; in a final labeling volume of 100 µl at 4°C. Higher concentrations of the label exceeded the maximum solubility and hence could not be used. Following labeling, the excess label was removed using a PD Minitrap G-25 desalting column (GE Healthcare). The labeled GyrA14-I491C was loaded onto a pre-equilibrated CcdB affinity column prepared using Affi-gel 10 (Bio-Rad) as per the manufacturer's instructions, and incubated at 4°C for 4 hours on a rotary mixer to enable binding. Unbound protein (containing FBS components) was removed by washing the column with 10 times the bed volume of coupling buffer (0.05 M sodium bicarbonate, 0.5 M NaCl, pH 8.3), and elution was done using 0.2 M Glycine, pH 2.5 into a vial containing an equal volume of 400 mM HEPES, pH 8.5. Mass spectra of the elutes were acquired on an UltrafleXtreme MALDI TOF (Bruker Daltonics). Mass of unlabeled test protein was 17.5 kDa, and upon labeling, an increase in mass of ~0.5 kDa was seen, corresponding to the mass of the Cys label. The approximate extent of Cys labeling of the GyrA14-I491C test protein was calculated as: $\%\ \text{Cys labeling} = \frac{\text{Peak intensity of labeled test protein}}{\text{Peak intensity of unlabeled test protein}} \times 100$

Cys labeling of virus.

Transient transfection of the pLAI-JRFL plasmid into HEK293T cells leads to production of replication competent WT LAI-JRFL virus. Prior to labeling, all virus stocks were made cell-free by filtration through a 0.45 µm syringe filter. 3×10^5 RLU equivalents (as determined by TZM-bl titrations) of WT virus was labeled with 1mM and 10 mM EZ-Link™ Maleimide-

PEG_2-Biotin (Thermo Fisher Scientific) in 1X PBS, overnight, in a reaction volume of 100 µl at 4°C. The effect of labeling on viral infectivity was measured in TZM-bl cells.

Neutralization potency (IC_{50}) of WT virus with bNAbs.

The neutralization potency (IC_{50}) of the WT LAI-JRFL virus with bNAbs VRC01, b12, PGT128 and PGT151 was estimated using a viral neutralization assay in TZM-bl cells. This assay measures the antibody-mediated neutralization of HIV-1, by quantifying the reduction in relative luminescence units (RLUs) i.e. Tat-driven luciferase gene expression upon viral infection. For this assay, 1-1.5x10^5 RLU equivalents of WT LAI-JRFL virus were incubated with serial threefold dilutions of bNAbs for 1 hour at 37°C in a total volume of 100 µl of 10% DMEM. To this, 1x10^4 freshly-trypsinized TZM-bl cells in 100µl containing 30µg/ml of DEAE-dextran were added, and incubated for 48 hours. Luminescence measurements were taken using the Britelite Plus reagent (PerkinElmer) on a VICTOR X2 luminometer (PerkinElmer). Neutralization potency (IC_{50}) is calculated as the sample dilution at which the RLUs in test wells (with bNAb) are reduced by 50% in comparison to RLUs in virus control wells (without bNAb) after subtraction of background RLUs from cell control wells. We used concentrations 100 times in excess of the calculated IC_{50} with different bNAbs[31, 32] for subsequent experiments.

Multiple cycle infectivity assay of Exposed Cys Library in HUT-R5 cells.

We carried out multi-cycle infectivity assays of the Exposed Cys Library in HUT-R5 cells (a derivative of the human T-cell line HUT-78, stably transduced with the CCR5 coreceptor). Multi-cycle infectivity assays evaluate both the infectivity as well as the rate and efficiency of virus production in cell culture over a course of time (replication fitness). For this assay, 5x10^6 HUT-R5 cells were infected with 3x10^5 RLU equivalents (as estimated from TZM-bl titrations) of the Exposed Cys Library using DEAE-dextran (at a final concentration of 30 µg/ml) in a

total volume of 250 µl of RPMI 1640 medium containing 10% FBS. Infection was carried out by incubation in a 37°C CO_2 incubator for 2 hours, with occasional shaking every half an hour. Excess virus from the infected cells was removed by washing with 45 ml of ice-cold 1X PBS. Infected cells were centrifuged at 1200 rpm for 5 mins, the supernatant was discarded, and the cell pellet was resuspended in 6 ml of 10% RPMI 1640, transferred to a T-25 flask and kept in a CO_2 incubator. Cells were split in a 1:2 ratio every alternate day post-infection, and virus replication was quantitated by measuring the p24 antigen released in culture supernatants at different time points post-infection using HIV-1 p24 SimpleStep ELISA kit® (Abcam). Serial dilutions of the HIV-1 p24 protein (strain 92418, subtype B) provided with the kit was used to plot a standard curve and replication fitness of the WT virus and Exposed Cys Library was measured by estimating the p24 levels from the standard curve.

Multi-cycle infectivity of labeled Exposed Cys Library in the presence of bNAbs.

To map epitopes of known bNAbs using our methodology, we carried out multi-cycle infectivity assays of the labeled Exposed Cys Library in the presence of 100 IC_{50} equivalents (and subsequent threefold dilutions) of bNAbs VRC01, PGT128 and PGT151. This was carried out by labeling the Exposed Cys Library, followed by incubation with bNAbs for 1 hour at 37°C, prior to carrying out multi-cycle infectivity assays in HUT-R5 cells.

Importantly, during the 1:2 split on every alternate day post-infection, additional bNAb was introduced into the culture media to keep the final bNAb concentration constant. Since epitopes for the selected bNAbs are present in the Exposed Cys Library, some mutant viruses are expected to be resistant to neutralization upon bNAb exposure. This was inferred by quantifying the ratio of p24 levels for the labeled Exposed Cys Library in the absence to presence of bNAb.

Neutralization titre (ID_{50}) determination of human plasma and guinea pig sera.

The neutralization titre or 50% inhibitory dose (ID_{50}) was determined using the TZM-bl cell-based viral neutralization assay. Neutralization assays was carried out for the WT LAI-JRFL virus in the presence of plasma from the elite neutralizer, NAB059 and also for terminal bleed sera (collected at week 34.5) from guinea pigs immunized with cyclically permuted gp120 trimers. Similar to IC_{50}, neutralization titre, or 50% inhibitory dose (ID_{50}) is calculated as the reciprocal of the serum dilution at which the RLUs in test wells (with plasma/sera) are reduced by 50% in comparison to RLUs in virus control wells (without plasma/sera) after subtraction of background RLUs from cell control wells. Titres 100 times in excess of the calculated ID_{50} were used to map epitopes of NAB059 plasma using the Exposed Cys Library. For mapping epitopes using the immunized guinea pig sera, 10 times the ID_{50} titre was used, since only limited amounts of these sera were available.

Multi-cycle infectivity assays of labeled Exposed Cys Library in the presence of plasma/sera.

To map epitopes present in polyclonal sera (Figure 1), multi-cycle infectivity assays in HUT-R5 cells[33] were carried out for the labeled Exposed Cys Library in the presence of 100 ID_{50} equivalents of NAB059 plasma or 10 ID_{50} equivalents of immunized guinea pig sera. Multi-cycle infectivity assays in the presence of plasma/sera were performed as described above for bNAbs. Importantly, during the 1:2 cell split on every alternate day post infection of the labeled library, additional plasma/sera was introduced into the culture media to keep the final titers constant. From the multi-cycle infectivity data, the ratio of p24 levels of the labeled Exposed Cys Library in the absence to presence of bNAb was also quantified.

Deep Sequencing of env gene for epitope mapping.

Replicative virus isolated from HUT-R5 culture supernatants on day 13 (post infection) was used to isolate RNA and prepare cDNA using RT-PCR. To this end, viral RNA was isolated from 500 µl of tissue culture supernatant harvested from HUT-R5 cells using a QIAamp® viral RNA Mini Kit (Qiagen). Elution was carried out using RNase-free 0.1% diethyl pyrocarbonate (DEPC)-treated water, instead of elution buffer AVE provided with the kit, since the latter contains 0.04% sodium-azide which affects A_{220} and A_{280} readings. RNA yield and purity were quantitated spectrophotometrically using NanoDrop™ 2000. SuperScript™ III reverse transcriptase (Thermo Fisher Scientific) was used to produce cDNA from viral RNA using a primer which binds 811 bp downstream of the *env* gene. The primer pLAI-JRFL-NefOR-rev has the sequence 5'-AGGCAAGCTTTATTGAGGC-3'. First-strand cDNA synthesis was carried out from the purified viral RNA in a 25 µl reaction containing 100 ng of viral RNA, 1 µl (200 units) of SuperScript™ III reverse transcriptase (Thermo Fisher Scientific), 5 µl of 5X First-Strand Buffer (Thermo Fisher Scientific), 0.5 µM of gene-specific primer (pLAI-JRFL-NefOR-rev), 1 µl of dNTP mix (containing 10 mM of each dNTP) and 1µl RNasin ribonuclease inhibitor (Promega). The cDNA synthesis reaction was carried out at 42°C for 50 mins, followed by inactivation of reverse transcriptase by heating at 70°C for 15 mins.

Subsequently, env was amplified using Phusion® High-Fidelity DNA Polymerase (NEB)using Phusion HF buffer and primers specific to the 5' and 3' ends of *env*, pLAI-JRFL-env-F (5'-TAGGCATCTCCTATGGCAGGAAG-3') and pLAI-JRFL-env-R (5'-GTCTCGAGATGCTGCTCCCACCCC-3'). PCR was done for 30 cycles with an annealing temperature of 60°C for 35 seconds. Amplified env cDNA was purified using a GeneJET PCR purification kit (Thermo Fisher Scientific). This step was necessary to remove the Phusion DNA polymerase which would otherwise degrade any overhangs subsequently generated.

This was further amplified for deep sequencing. The length of the env gene is ~2.2 kb, whereas the maximum length of DNA that can be sequenced on the Illumina HiSeq 2500 platform is 500 bases in a paired end sequencing run (2x250 bp). Because of the size restriction imposed by Illumina sequencing, the env gene was sequenced in six fragments.

The env gene was sequenced from the Exposed Cys Library isolated under different conditions (listed in Table S3). Primers were designed for each of the conditions, with a unique Multiplex IDentifier (MID) tag for each condition. For the first 5 fragments (Figure S16) both the forward and reverse flanking primers begin with NNN, followed by 6 bases of the unique sequence tag for each MID and the 21 bases complementary to the gene. The sixth fragment contained the same flanking forward primer, but the reverse primer was devoid of the MID sequence. Each MID sequence corresponds to one of the following conditions: 3 broadly neutralizing antibodies (bNAbs VRC01, PGT128, PGT151), 2 sera (NAB059 patient sera and trimeric cyclic permutant (CycP) guinea pig sera) and 3 controls (conditions 6-8, Table S3). The PCR products have an average read length of 420 bp, with an overlap of ~63 bp between successive fragments. PCR for the fragments was carried out using Phusion polymerase with a template concentration of 10ng/µl for 15 cycles. PCR product concentration following agarose gel electrophoresis was quantified by densitometry analysis using the Bio-Rad Quantity One software (Bio-Rad). Fragments were pooled in an equimolar ratio, gel purified and sequenced on the Illumina HiSeq 2500 platform.

Deep Sequencing Data Analysis.

The script for the deep sequencing data analysis was developed using PERL with certain modifications from an already existing in-house protocol (https://github.com/skshrutikhare/cys_library_analysis). The raw data consisting of 134897599 x 2 paired reads of the mutants were assembled using PEAR v0.9.6 (Paired-End

Read Merger) tool [34]. Out of this 1.102% reads were discarded and the remaining 133410630 assembled reads were obtained. Initial screening for the env reads was based on the unique primer pairs for each fragment and were done separately. The same set of MIDs were used for each fragment. The last fragment had MID at only the 5' end, so scripts were slightly modified, but the remaining five fragments were processed similarly. A Phred score cutoff of ≥ 20 was applied on these reads for quality filtering. After the quality filtering, reads not having the correct MID/primer set for env data were eliminated along with the reads having mismatched MIDs. Each filtered read will thus have only the MID, forward/reverse primer along with the gene segment of interest. Reads in each fragment were segregated into different bins based on the MID. Before binning, reads with insertions/deletions, with incomplete primer/MIDs (due to quality filtering) and those having incorrect pairs of primers were removed. Binned reads were aligned with the wild type sequence of the env gene corresponding to each fragment using Water v6.4.0.0 program[35]. Aligned reads were reformatted for ease of further processing. Finally, these reads were classified based on whether they were wildtype, or contained indels and substitutions (single, double, triple and multiple mutants). Single Cys mutants were sorted based on read numbers. Normalization was carried out as follows: for Residue i in Fragment x at MID y, where x = 1 – 6, and y = 1 - 5

Numerator (N) = (# Cys reads (i) / (Total (#Cys reads) + WT reads)) for Fragment x at MID y

Denominator (D) = (# Cys reads (i) / Total (#Cys reads)+ WT reads)) for Fragment x at MID 6. Normalized Value (R) = N/D. All residues with R > 0.5 were designated as epitope residues.

Alternative methods of normalization, for example, using the read ratio of a given Cys mutant at MID x, where x = 1-5 to that at MID 6 yielded very similar conclusions.

Titration of viral supernatant using qRT-PCR of the *pol* gene.

Viral titre (or viral load) was measured using quantitative real time-PCR (qRT-PCR) of the conserved HIV-1 *pol* gene (201 bp) using SYBR Green detection chemistry. To this end, serial dilutions of the plasmid pLAI-JRFL were prepared, ranging from 10^1 to 10^8 copies/µl. Dilutions of the plasmid were prepared in nuclease-free water containing 0.1 µg/µl BSA. qRT-PCR reactions were performed in 20 µl reaction volumes containing 10 µl of 2X iQ™ SYBR® Green Supermix (Bio-Rad), 0.5 µM each of *pol* gene-specific forward and reverse primers, 1µl of the serially diluted plasmid DNA template and 2 µg of BSA. Volume was adjusted using nuclease-free water. A no template control reaction was also setup which contained all the reaction components except the template. Sequences of the primers used to amplify the *pol* gene: pLAI-JRFL-pol-forward: 5'-AGGAGCAGAAACGTTCTATGTAGATGG-3' and pLAI-JRFL-pol-reverse: 5'-GCATATTGTGAGTCTGTTACTATATTTACT-3'. The qRT-PCR assay was performed on an iQ5 RT-PCR detection system (Bio-Rad) using the following cycling conditions: initial denaturation at 94°C for 5 mins, followed by 40 cycles of denaturation at 94°C for 30 secs, primer annealing at 60°C for 35 secs, and primer extension at 72°C for 30 secs. The threshold cycle (C_T) values were obtained from the fluorescence data during amplification and used to plot a standard curve for the different dilutions of pLAI-JRFL plasmid DNA. Melting curve analysis and agarose gel electrophoresis of the qPCR products were also carried out to confirm that the observed amplification was specific.

The standard curve was used for the absolute quantification of viral cDNA. For this, viral RNA was isolated from 500 µl of tissue culture supernatant harvested from HEK293T cells using a QIAamp® viral RNA Mini Kit (Qiagen), following the manufacturer's instructions. RNA yield and purity were quantitated spectrophotometrically using NanoDrop™ 2000. First-strand cDNA synthesis was carried out from the purified viral RNA in a 25 µl reaction containing 100 ng of viral RNA, 1 µl (200 units) of SuperScript™ III reverse transcriptase (Thermo Fisher

Scientific), 5 µl of 5X First-Strand Buffer (Thermo Fisher Scientific), 0.5 µM of *pol* gene-specific primer (pLAI-JRFL-pol-reverse), 1 µl of dNTP mix (containing 10 mM of each dNTP) and 1µl RNasin ribonuclease inhibitor (Promega). The cDNA synthesis reaction was carried out at 42°C for 50 mins, followed by inactivation of reverse transcriptase by heating at 70°C for 15 mins.

To estimate viral titres, 2µl of viral cDNA from the first-strand reaction was used for absolute quantification using the standard curve. For this, qRT-PCR of the *pol* gene was carried out using SYBR Green chemistry using the same reaction components and cycling conditions that had been used earlier for plasmid DNA, except that viral cDNA was used as the template. The C_T value obtained from the qRT-PCR data was used to estimate viral copy numbers from the standard curve. An important assumption of this method is that the reverse transcription step is 100% efficient. Although the qRT-PCR assay is capable of accurately measuring viral titres, it does not accurately measure infectious titres, since this assay also accounts for defective particles that harbour non-native forms of the Env and consequently, do not give rise to productive infection.

Determination of infectious viral titre using TZM-bl cells.

Infectious titre from HEK293T viral supernatants was estimated using a luciferase reporter-based assay in TZM-bl cells. Prior to labelling, all virus stocks were made cell-free by filtration through a 0.45 µm syringe filter. $3x10^5$ RLU equivalents (as determined by TZM-bl titrations) of WT and Cys-mutant viruses were labelled with 10 mM EZ-Link™ Maleimide-PEG_2-Biotin (Thermo Fisher Scientific), overnight in a reaction volume of 100 µl at 4°C. Using the viral titres obtained from the qRT-PCR assay, $3x10^5$ virions (and subsequent threefold dilutions) were used to infect $5x10^5$ TZM-bl cells, at a resultant multiplicity of infection of 0.6 (for the

undiluted virus), and infectivity of the WT virus and Cys-mutant viruses was quantified by luminescence measurements from the cells 48 hours post infection.

Acknowledgements:

The authors would like to thank Dr. John R Mascola, Dr. Dennis Burton, the NIH AIDS reagent program and the IAVI Neutralizing Antibody Consortium for providing HIV antibodies and reagents that enabled the study. The authors are also grateful to Jyothi Prabha Goli for providing purified GyrA14-Cys protein for experiments and Dr. Shruti Khare for running PERL scripts to calculate pairwise sidechain centroid distances. Dr. Ujjwal Rathore and Dr. Shahbaz Ahmed is acknowledged for helpful discussions.

Author Contributions:

RD and RV designed the study, interpreted and analyzed data, and wrote the manuscript. RD performed experiments. RRC helped with cloning Cys mutants and performed PCRs of the env gene for deep sequencing. KM processed deep sequencing data. LEH provided reagents which enabled the study.

Competing Interests Statement:

RD and RV are inventors of a patent on mapping protein binding sites and conformational epitopes using Cys labeling (PCT WO/ 2018/104967).

References:

1. Najar, T.A., Khare, S. & Varadarajan, R. Rapid Mapping of Protein Binding Sites and Conformational Epitopes by Coupling Yeast Surface Display to Chemical Labeling and Deep Sequencing *Methods Mol Biol* **1785**, 77-88 (2018).
2. Najar, T.A., Khare, S., Pandey, R., Gupta, S.K. & Varadarajan, R. Mapping Protein Binding Sites and Conformational Epitopes Using Cysteine Labeling and Yeast Surface Display *Structure* **25**, 395-406 (2017).
3. Sahoo, A., Khare, S., Devanarayanan, S., Jain, P.C. & Varadarajan, R. Residue proximity information and protein model discrimination using saturation-suppressor mutagenesis *Elife* **4** (2015).

4. Cheedarla, N. et al. Evolution of Neutralization Response in HIV-1 Subtype C-Infected Individuals Exhibiting Broad Cross-Clade Neutralization of HIV-1 Strains *Front Immunol* **9**, 618 (2018).
5. Cheedarla, N. et al. Broad and potent cross clade neutralizing antibodies with multiple specificities in the plasma of HIV-1 subtype C infected individuals *Sci Rep* **7**, 46557 (2017).
6. Saha, P. et al. Designed cyclic permutants of HIV-1 gp120: implications for envelope trimer structure and immunogen design *Biochemistry* **51**, 1836-1847 (2012).
7. Kesavardhana, S. et al. Structure-based Design of Cyclically Permuted HIV-1 gp120 Trimers That Elicit Neutralizing Antibodies *J Biol Chem* **292**, 278-291 (2017).
8. Jones, A.T. et al. HIV-1 vaccination by needle-free oral injection induces strong mucosal immunity and protects against SHIV challenge *Nat Commun* **10**, 798 (2019).
9. Jones, A.T. et al. A Trimeric HIV-1 Envelope gp120 Immunogen Induces Potent and Broad Anti-V1V2 Loop Antibodies against HIV-1 in Rabbits and Rhesus Macaques *J Virol* **92** (2018).
10. Tripathi, A. & Varadarajan, R. Residue specific contributions to stability and activity inferred from saturation mutagenesis and deep sequencing *Curr Opin Struct Biol* **24**, 63-71 (2014).
11. Heydenreich, F.M. et al. High-throughput mutagenesis using a two-fragment PCR approach *Sci Rep* **7**, 6787 (2017).
12. Dingens, A.S., Arenz, D., Weight, H., Overbaugh, J. & Bloom, J.D. An Antigenic Atlas of HIV-1 Escape from Broadly Neutralizing Antibodies Distinguishes Functional and Structural Epitopes *Immunity* **50**, 520-532 e523 (2019).
13. Bloom, J.D. An experimentally determined evolutionary model dramatically improves phylogenetic fit *Mol Biol Evol* **31**, 1956-1978 (2014).
14. Frillingos, S., Sahin-Toth, M., Wu, J. & Kaback, H.R. Cys-scanning mutagenesis: a novel approach to structure function relationships in polytopic membrane proteins *FASEB J* **12**, 1281-1299 (1998).
15. Zhu, Q. & Casey, J.R. Topology of transmembrane proteins by scanning cysteine accessibility mutagenesis methodology *Methods* **41**, 439-450 (2007).
16. Lee, P.A. et al. Cysteine-scanning mutagenesis and disulfide mapping studies of the conserved domain of the twin-arginine translocase TatB component *J Biol Chem* **281**, 34072-34085 (2006).
17. Mueckler, M. & Makepeace, C. Cysteine-scanning mutagenesis and substituted cysteine accessibility analysis of transmembrane segment 4 of the Glut1 glucose transporter *J Biol Chem* **280**, 39562-39568 (2005).
18. Chalker, J.M., Bernardes, G.J., Lin, Y.A. & Davis, B.G. Chemical modification of proteins at cysteine: opportunities in chemistry and biology *Chem Asian J* **4**, 630-640 (2009).
19. Kim, Y. et al. Efficient site-specific labeling of proteins via cysteines *Bioconjug Chem* **19**, 786-791 (2008).
20. Yuan, Y. et al. Labeling thiols on proteins, living cells, and tissues with enhanced emission induced by FRET *Sci Rep* **3**, 3523 (2013).
21. Bianchi, M. et al. Electron-Microscopy-Based Epitope Mapping Defines Specificities of Polyclonal Antibodies Elicited during HIV-1 BG505 Envelope Trimer Immunization *Immunity* **49**, 288-300 e288 (2018).
22. Bajaj, K., Chakrabarti, P. & Varadarajan, R. Mutagenesis-based definitions and probes of residue burial in proteins *Proc Natl Acad Sci U S A* **102**, 16221-16226 (2005).
23. Mikell, I. et al. Characteristics of the earliest cross-neutralizing antibody response to HIV-1 *PLoS Pathog* **7**, e1001251 (2011).
24. Moore, P.L. et al. Potent and broad neutralization of HIV-1 subtype C by plasma antibodies targeting a quaternary epitope including residues in the V2 loop *J Virol* **85**, 3128-3141 (2011).
25. Permar, S.R. et al. Maternal HIV-1 envelope-specific antibody responses and reduced risk of perinatal transmission *J Clin Invest* **125**, 2702-2706 (2015).
26. Scharf, L. et al. Structural basis for germline antibody recognition of HIV-1 immunogens *Elife* **5** (2016).
27. Weidenbacher, P.A. & Kim, P.S. Protect, modify, deprotect (PMD): A strategy for creating vaccines to elicit antibodies targeting a specific epitope *Proc Natl Acad Sci U S A* **116**, 9947-9952 (2019).

28. Jain, P.C. & Varadarajan, R. A rapid, efficient, and economical inverse polymerase chain reaction-based method for generating a site saturation mutant library *Anal Biochem* **449**, 90-98 (2014).
29. Baliga, C., Varadarajan, R. & Aghera, N. Homodimeric Escherichia coli Toxin CcdB (Controller of Cell Division or Death B Protein) Folds via Parallel Pathways *Biochemistry* **55**, 6019-6031 (2016).
30. Spicer, C.D. & Davis, B.G. Selective chemical protein modification *Nat Commun* **5**, 4740 (2014).
31. Das, S.R. et al. Defining influenza A virus hemagglutinin antigenic drift by sequential monoclonal antibody selection *Cell Host Microbe* **13**, 314-323 (2013).
32. Doud, M.B., Lee, J.M. & Bloom, J.D. How single mutations affect viral escape from broad and narrow antibodies to H1 influenza hemagglutinin *Nat Commun* **9**, 1386 (2018).
33. Ghimire, D., Timilsina, U., Srivastava, T.P. & Gaur, R. Insights into the activity of maturation inhibitor PF-46396 on HIV-1 clade C *Sci Rep* **7**, 43711 (2017).
34. Zhang, J., Kobert, K., Flouri, T. & Stamatakis, A. PEAR: a fast and accurate Illumina Paired-End reAd mergeR *Bioinformatics* **30**, 614-620 (2014).
35. Smith, T.F. & Waterman, M.S. Identification of common molecular subsequences *J Mol Biol* **147**, 195-197 (1981).

Figure Legends:

Figure 1. Strategy for rapid epitope mapping of bNAbs against HIV-1.

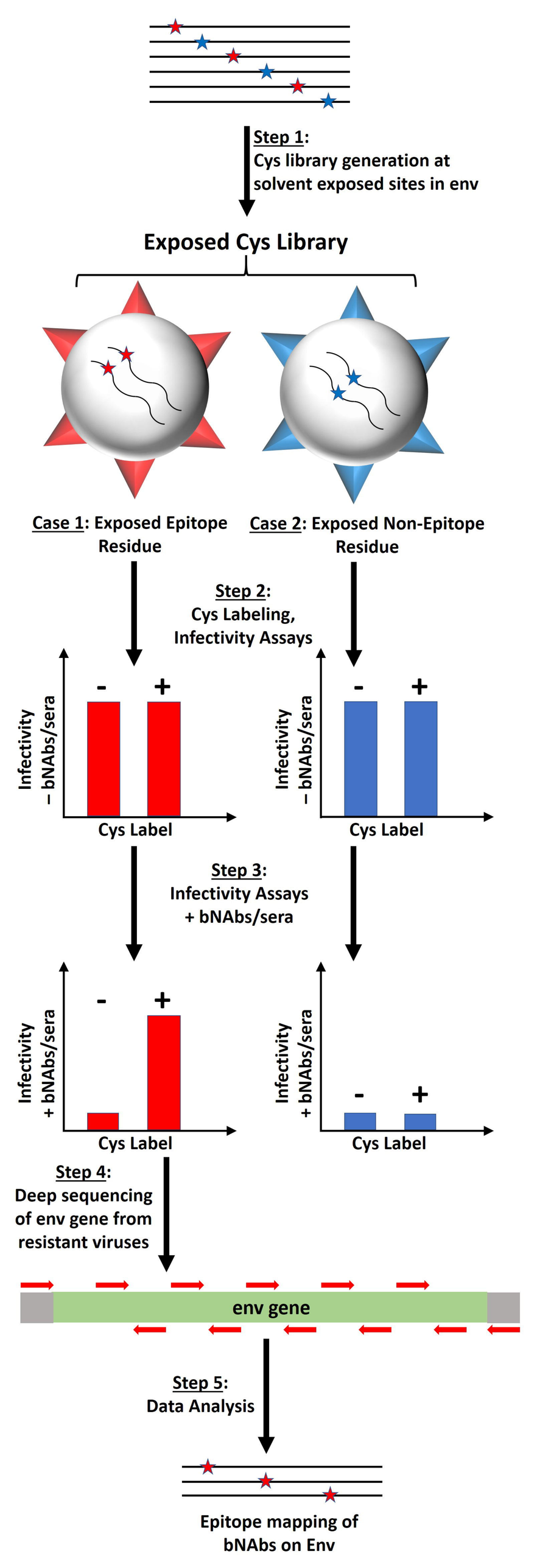
Step 1: A library of single Cys mutants was created using site-directed mutagenesis at selected, solvent exposed sites in the env gene of the HIV-1 molecular clone, pLAI-JRFL. The Env ectodomain in this molecular clone is from the JRFL strain (residues 25 – 690), whereas the rest of the sequence is derived from the LAI strain of HIV-1. An equimolar pool of single-site Cys mutants of *env* gene (cloned in pLAI-JRFL) was used to transfect HEK293T cells. This led to the production of a library of mutant viruses, having Cys mutations at the selected solvent exposed residues in the Env ectodomain; henceforth referred to as the Exposed Cys Library. Step 2: The pool of mutant viruses was covalently labeled with a Cys reactive reagent (CRR), EZ-Link™ Maleimide-PEG$_2$-Biotin (Thermo Fisher Scientific), and infectivity was monitored before and after labeling. Step 3: At epitope residues, masking by the bulky Cys label prevents bNAbs from binding and neutralizing the epitope mutant viruses. Steps 4 & 5: Deep sequencing of the env gene from antibody/sera-resistant virion pools was used to identify epitope residues at single residue resolution.

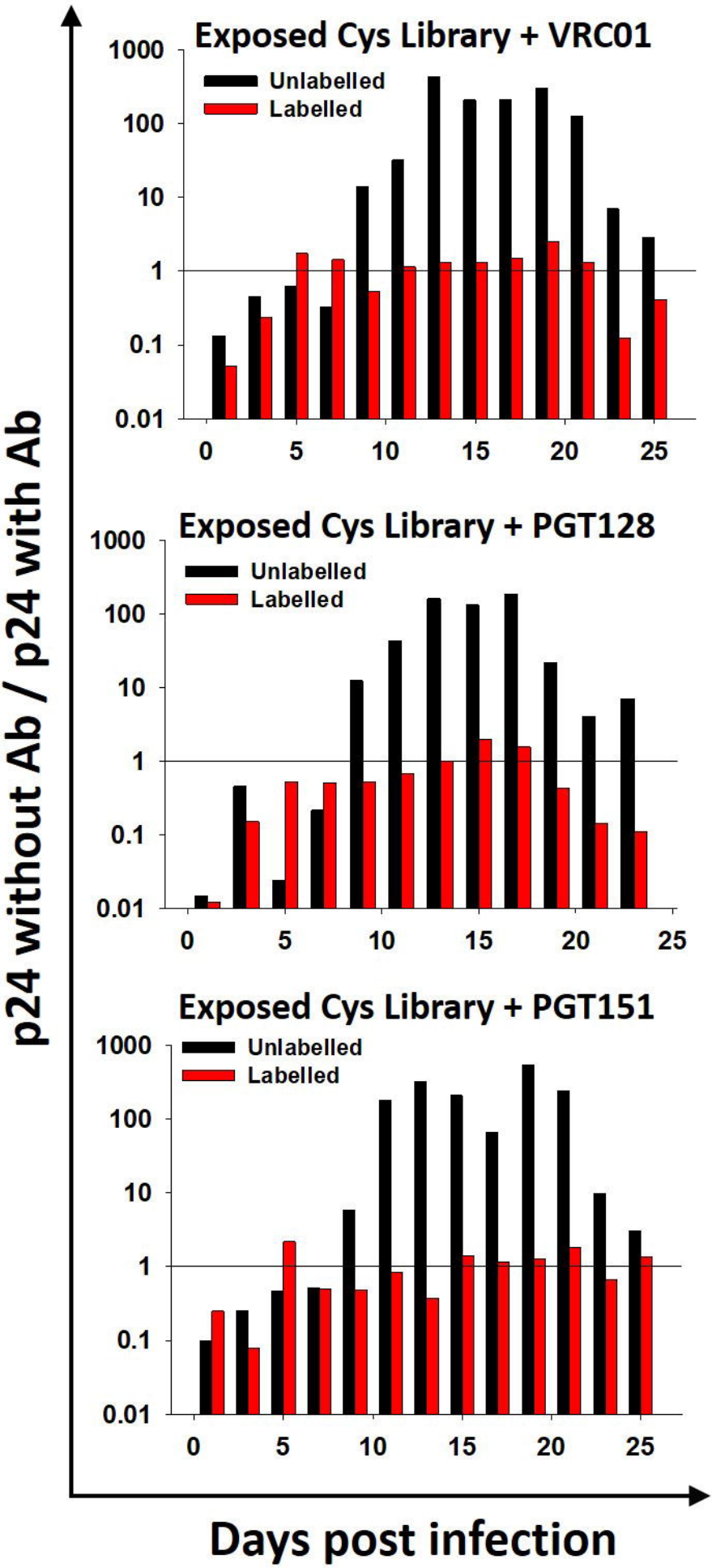
Figure 2. Ratio of p24 levels for the labeled Exposed Cys Library in the absence to presence of bNAbs.

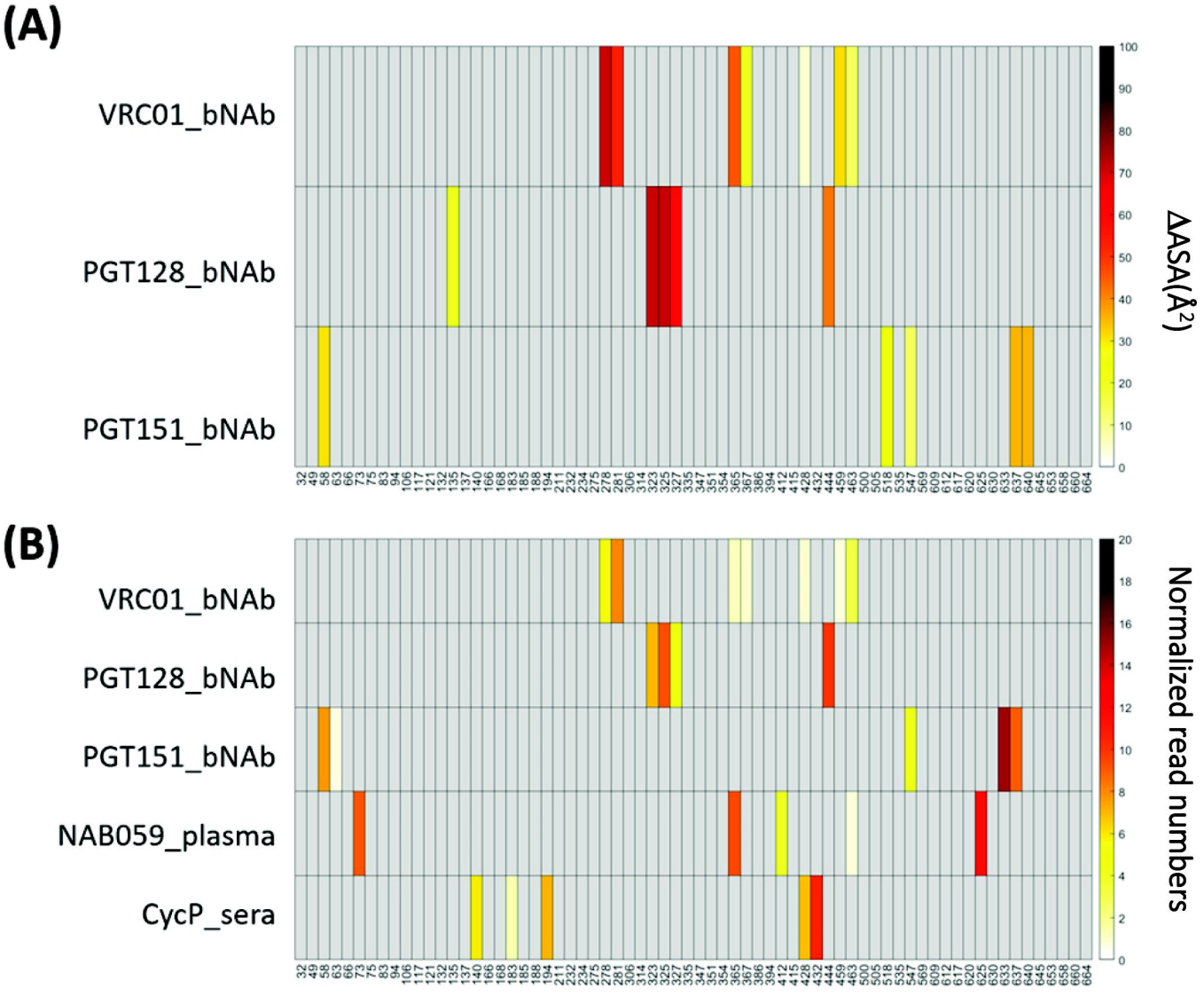
The ratio of p24 levels was estimated from multi-cycle infectivity assays in the presence of bNAbs VRC01, PGT128 and PGT151. After ~10 days of infection, the ratio of p24 levels of the labeled viral library in the absence to presence of bNAbs is ~1 (red bar), confirming that labelling at exposed, epitope residues on Env prevents neutralization by the corresponding bNAb. In contrast, in the absence of labelling, viral entry is abolished in the presence of the bNAb (black bar), leading to a high value of the ratio.

Figure 3. Epitope mapping of bNAbs and sera/plasma. (A) Change in accessible surface area upon binding to selected bNAbs (B) Normalized read frequencies following deep sequencing.

The solvent accessible surface area for residue sidechains was calculated using the NACCESS V 2.1.1 program. The buried surface area is calculated by subtracting the accessible surface area of Env residue sidechains in the presence of the bNAb from the area in the absence of the bNAb (ΔASA). Surface areas buried in complexes of Env with bNAbs VRC01, PGT128 and PGT151 (PDB IDs 3NGB (for VRC01), 5C7K (for PGT128) and 5FUU (for PGT151) respectively) were calculated. Read numbers for selected Cys mutants for the different selection conditions (MIDs 1-5, corresponding to bNAbs and polyclonal sera) were obtained from the deep sequencing data and normalized by the reads in the unselected condition, in the absence of bNAbs (MID 6), as described in the Methods section.


Step 1:
Cys library generation at solvent exposed sites in env
Exposed Cys Library
Case 1: Exposed Epitope Residue
Case 2: Exposed Non-Epitope Residue
Step 2:
Cys Labeling, Infectivity Assays
Infectivity – bNAbs/sera
-
+
Cys Label
Infectivity – bNAbs/sera
-
+
Cys Label
Step 3:
Infectivity Assays + bNAbs/sera
Infectivity + bNAbs/sera
-
+
Cys Label
Infectivity + bNAbs/sera
-
+
Cys Label
Step 4:
Deep sequencing of env gene from resistant viruses
env gene
Step 5:
Data Analysis
Epitope mapping of bNAbs on Env


Exposed Cys Library + VRC01
Unlabelled
Labelled
1000
100
10
1
0.1
0.01
0
5
10
15
20
25
Exposed Cys Library + PGT128
Unlabelled
Labelled
1000
100
10
1
0.1
0.01
0
5
10
15
20
25
Exposed Cys Library + PGT151
Unlabelled
Labelled
1000
100
10
1
0.1
0.01
0
5
10
15
20
25
p24 without Ab / p24 with Ab
Days post infection


(A)
VRC01_bNAb
PGT128_bNAb
PGT151_bNAb
ΔASA(Å²)
(B)
VRC01_bNAb
PGT128_bNAb
PGT151_bNAb
NAB059_plasma
CycP_sera
Normalized read numbers