Rationally designed carbohydrate-occluded epitopes elicit HIV-1 Env-specific antibodies

Cheng Zhu¹,², Elena Dukhovlinova³, Olivia Council³,⁴, Lihua Ping³, Edgar M. Faison¹, Shamit S. Prabhu⁵, E. Lake Potter³, Stephen L. Upton¹, Guowei Yin¹, James M. Fay¹, Laura P. Kincer³,⁴, Ean Spielvogel³, Sharon L. Campbell¹,³, S. Rahima Benhabbour⁵,⁶, Hengming Ke¹,³, Ronald Swanstrom¹,³ & Nikolay V. Dokholyan¹,²,³,⁶,⁷

An array of carbohydrates masks the HIV-1 surface protein Env, contributing to the evasion of humoral immunity. In most HIV-1 isolates ‘glycan holes’ occur due to natural sequence variation, potentially revealing the underlying protein surface to the immune system. Here we computationally design epitopes that mimic such surface features (carbohydrate-occluded neutralization epitopes or CONE) of Env through ‘epitope transplantation’, in which the target region is presented on a carrier protein scaffold with preserved structural properties. Scaffolds displaying the four CONEs are examined for structure and immunogenicity. Crystal structures of two designed proteins reflect the computational models and accurately mimic the native conformations of CONEs. The sera from rabbits immunized with several CONE immunogens display Env binding activity. Our method determines essential structural elements for targets of protective antibodies. The ability to design immunogens with high mimicry to viral proteins also makes possible the exploration of new templates for vaccine development.
**Results**

**Structure-based immunogen design.** Atomic-level details of the fully glycosylated, pre-fusion conformation of the Env trimer have been obtained as crystal and cryo-electron microscopy structures\(^2,22\). The conformations of the CONEs are well conserved in the structures of Env from clades A, B, C, and G (Supplementary Fig. 1b). We identified the substructural elements of each CONE based on the BG505 S001.664 trimer structure (Protein Data Bank [PDB]: 4TVP and 4NCO) and adopted the consensus sequence of clade C HIV-1 Env as the reference for the immunogen design\(^23,24\) (Fig. 1d and Supplementary Fig. 2). The rationale for identifying CONEs include: (i) the change of solvent-exposed area on the Env surface due to absence of glycans. The boundaries of CONE residues can be determined by modeling the interactions between underglycosylated Env and immunoglobulin (Supplementary Fig. 3). (ii) whether the substructures of CONEs can be readily matched to existing scaffolds in PDB (see Methods). (iii) conservation of surface glycans\(^10,11\). Using a rapid structural motif-mining algorithm Erebus\(^25\) (Supplementary Fig. 4, workflow), we searched the PDB for scaffolds capable of supporting the disembodied CONE structures (three-stranded \(\beta\)-sheet, a-helix, and two separate loops). We estimated the surface matching between scaffolds and CONEs by analysis of root-mean-square deviations (RMSD, 0.5–3.0 Å) and selected the top-ranking scaffolds for each CONE (Supplementary Table 1). The surface amino acids within the CONE-like region within each scaffold were replaced by solvent-exposed residues in the consensus sequences of CONEs (Supplementary Figs 2 and 4). For instance, a discontinuous three \(\beta\)-strand motif was grafted onto CONE 1 scaffolds, while the two helices on the CONE 2 and scaffold 5 design (C2S5, a similar nomenclature is followed for all designed proteins) displayed the \(\alpha\)2 helix-specific residues (Supplementary Fig. 5). The modified scaffold was subsequently relaxed by discrete molecular dynamics\(^26\).

For each scaffold, multiple designs (86 in total) were generated adopting various combinations of CONE residues and their substitutes in the consensus sequences. We then estimated the change in free energy of these substitutions (\(\Delta\Delta \text{G}_{\text{mut}}\)) using the Eris molecular design suite and discarded highly destabilizing mutations (55 designs of \(\Delta\Delta \text{G}_{\text{mut}} > 6 \text{kcal mol}^{-1}\))\(^27\). The conformational stabilities of all CONE mimetics were further evaluated through molecular dynamics simulations, which allowed us to rapidly screen the designed protein library and choose the scaffolds that exhibited substantial rigidity around the CONE segment (root-mean-square fluctuations (RMSF) analysis, Supplementary Fig. 6). A recent study of massively parallel protein design demonstrated that structural elements of the most successful designs exhibit high similarity to natural proteins of similar local sequences\(^28\). Hence, we reasoned that the success of epitope transplantation would correlate with the compatibility between local structure of grafted sequences and their native conformation in Env, whereas higher RMSF of grafted residues indicates that the design explores wider conformational states and likely displays poor agreement to the target structure. Therefore, we filtered out the designs featuring significant plasticity of the grafted epitopes or designated them as alternative scaffolds (21 designs of RMSF > 3.5 Å at CONEs). Overall, three to five designs of each CONE were expressed in Escherichia coli (Supplementary Fig. 5). Of the sixteen candidates tested (one most stable design for each scaffold), eight could be successfully purified as soluble proteins (Supplementary Table 2).

**Biophysical characterization of designed proteins.** The eight proteins exhibited circular dichroism (CD) spectra consistent with the designed topology (Supplementary Fig. 7a). The stabilities of the designs to thermal denaturation were assessed by CD spectroscopy. Their melting temperatures (\(T_{\text{m}}\)) were in the range of 42–75°C (Supplementary Fig. 7b). The CONE 1, 2, and 4 proteins were monomeric in solution (Supplementary Fig. 7c, d), while the CS3 design displayed rapid exchange between monomeric and dimeric states. We selected the most stable designs
(C1S1, C2S5, and C4S3) for immunization experiments and further structural analysis.

To determine whether the designs recapitulate the native conformations of each CONE, we solved the crystal structures of C2S5 (α2 helix) and C4S3 (loop C) at resolutions of 2.0 Å and 1.2 Å, respectively (PDB 6CFE and 6CBU, Table 1). The electron density maps revealed that the side chains of the grafted residues were well positioned relative to the computational model (Supplementary Fig. 8). Comparison with the Env trimer protein revealed a high degree of mimicry: within the epitope region, the Ca RMSDs between the designed proteins and Env were 0.42 Å (C2S5) and 0.34 Å (C4S3), respectively, suggesting that our CONE mimetics accurately display the Env residues (Fig. 2a, b).

The C2S5 scaffold is built on a four-helix bundle from apolipoprotein E3. The two helixes used to display the CONE 2 residues were well matched to the α2 helix in terms of the Ca-Cβ distances.
immunized small groups of rabbits (four to six per group) with the designed immunogens to raise CONE-specific antibodies. Two different immunization strategies were applied: incomplete Freund’s adjuvant (FA) and a nanoparticle (NP) formulation without adjuvant (Supplementary Tables 3, 4, Supplementary Fig. 10). In these initial immunization experiments, we sought a qualitative understanding of the nature of the antibodies that can be generated to each CONE. The CONE 1, 2, 4, and 5 mimetics were all immunogenic, giving rise to autologous binding antibodies as determined by enzyme-linked immunosorbent assay (ELISA, Fig. 3a). The FA-associated immunization protocols induced antibodies with the titers more than 10-fold higher than those induced by NP-associated immunization. As a control experiment, we also immunized five rabbits with a synthetic peptide of the CONE 4 sequence (LAEEEL). None of these rabbits could induce antisera of significant titters (Supplementary Table 5), in contrast to the C4S3-immunized rabbits, indicating that for a linear epitope, constraining it in the right conformation with suitable protein scaffolds is favorable for the generation of functional antibodies.

We then evaluated antibody specificity for the CONE-derived epitopes (Supplementary Fig. 11). For each immunogen, an alternative design (C1S5, C2S3, C4S2, and C5S2, Supplementary Table 2) with the same CONE grafted onto a different scaffold protein was used as a control to HIV-specific responses. No HIV-specific antibodies were obtained after immunization with CONE 2 (a2 helix). The CONE 4 (loop C) and CONE 5 (loop E) immunogens did induce HIV-specific antibodies as assessed by ELISA and western blot assay. The purified IgG fractions of anti-CONE 4 and anti-CONE 5 rabbit sera bound to the deglycosylated form of gp120 but not the glycosylated protein (Fig. 3b).

We also analyzed the anti-Env responses of individual rabbit sera in an ELISA assay using the deglycosylated Env SOSIP trimer. We modified the sequence of SOSIP from the original BG505 SOSIP.664 isolate23, so that the exposed residues at the CONE regions represent the consensus sequences of Env subtype C. The sera of the rabbits immunized with C1S1 (FA protocol) contained antibodies that bound Env, featuring a significantly elevated response to Env in comparison to the homologous pre-immune sera; in contrast, only one animal generated strong-binding antibodies with C1S1-NPs (Supplementary Fig. 12). Most of the animals immunized with C4S3 also generated antibodies that bound to the deglycosylated SOSIP Env protein, whereas no such antibodies were detected in animals immunized with C5S3.

Four C1S1-immunized rabbits generated sera with the highest reactivity to the SOSIP trimer. We tested these serum samples in a virus neutralization assay against reporter HIV-1 viruses, which were pseudotyped using three different env gene clones (682, 1086, and 3037, all subtype C). The viruses expressing the WT Env represent the samples isolated from infected subjects, while the fully glycosylated Env and the mutants represent the CONE-concealed and CONE-revealed samples, respectively (Fig. 4). In this analysis, we accounted for the animal-to-animal variation by comparing responses of the pre-immune serum and the immunization endpoint serum for each rabbit and each virus. The four post-immunization rabbit sera exhibited modest virus inhibitory activity in comparison to the homologous pre-immune serum. The highest responses we observed (50–70% inhibition) were with virus 3037 (missing N332 glycan) by the serum of rabbit 481, and with virus 1086 (missing N332 glycan) by the serum of rabbit S1206. Given the low level of virus inhibition associated with these sera it is difficult to distinguish if this activity represents true virus neutralization or some type of nonspecific inhibition of infectivity, although we note this activity is differentially present after immunization. The observed inhibitory activity did not depend critically on the absence of the CONE 1-associated glycans on the Env protein surface. If this orientations (Fig. 2a and Supplementary Fig. 8d). The CONE 4 (loop C) residues were grafted onto the C-terminal loop region of an α/β mixed protein (Fig. 2b). A characteristic array of three glutamic acids flanked by hydrophobic residues defines the conformation of CONE 4 loop in Env. Our design largely recapitulates the conformations of two charged side chains (E88 and E89). The hydrophobic residue L86 also adopts a similar conformation of CONE 4 loop in Env. Our design largely recapitulates the conformations of two charged side chains (E88 and E89). The hydrophobic residue L86 also adopts a similar

### Table 1 Crystallographic data collection and refinement statistics

|             | C2S5       | C4S3       |
|-------------|------------|------------|
| **Data collection** |            |            |
| Space group  | P3 21      | P2 2 1     |
| **Cell dimensions** |            |            |
| a, b, c (Å) | 46.88, 46.88, 140.42 | 26.07, 57.61, 62.94 |
| α, β, γ (°) | 90, 90, 120 | 90, 90, 90 |
| Resolution (Å) | 50.00–2.00 | 50.00–1.20 |
| Rmerge (%)  | (2.03–2.00) | (1.22–1.20) |
| Rwork/Rfree (%) | 0.064 (0.886) | 0.056 (0.130) |
| I/σ(I)      | 13.6 (3.7) | 18.9 (6.9) |
| Completeness (%) | 99.67 | 98.40 |
| Redundancy | 18.0 (16.5) | 14.1 (13.9) |
| No. of unique reflections | 12904 | 30098 |
| **Refinement statistics** |            |            |
| Resolution (Å) | 39.00–2.00 | 42.49–1.20 |
| No. of reflections | 12849 | 28610 |
| Rwork/Rfree (%) | 23.65/26.87 | 14.30/15.30 |
| No. of atoms |            |            |
| Protein atoms | 2344 | 771 |
| Ligand/ion | 0 | 1 (SO42−) |
| Water | 17 | 67 |
| Average B-factor (Å²) | 70.43 | 7.54 |
| Protein | | |
| Ligand/ion | NA | 7.69 |
| Water | 48.78 | 14.10 |
| **RMS deviation from ideality** |            |            |
| Bond lengths (Å) | 0.004 | 0.006 |
| Bond angles (°) | 0.513 | 1.198 |
| Ramachandran statistics |            |            |
| Favored regions (%) | 99.28 | 89.20 |
| Allowed regions (%) | 0.72 | 10.80 |
| Outliers (%) | 0 | 0 |

*Highest resolution shell statistics are shown in parentheses

**As defined by MolProbity**
does represent antibody-mediated neutralization, then the rabbit-raised CONE 1-specific antibodies may possess long loops in the complementarity-determining region that are capable of penetrating the glycan shield, as seen in various neutralizing antibodies characterized previously. In this scenario the occluded surface features may not be immunogenic under the glycan shield, but the corresponding antibodies when they exist may be able to intercalate between structurally dynamic glycan side chains, thus resulting in neutralization.

**Discussion**

The strategies that HIV-1 adopts to avoid in vivo neutralization, including hypervariable protein sequences and a glycan coat that occludes the conserved protein surface, confounded the attempts at immunization using Env trimer as antigens. Here we described the implementation of an approach based on the observation that less conserved glycosylation sites are occasionally absent due to the intrinsic sequence variation in HIV-1. Our results point to the possibility of generating antibodies that target surface features of Env at ‘glycan holes’. We found that binding of anti-CONEs 4 and 5 antibodies to denatured gp120 critically depended on the removal of the proximal glycan. Also, the reactivities of anti-CONEs 1 and 4 antibodies with the SOSIP Env were greatly enhanced after deglycosylation. The magnitude of inhibitory effect upon HIV-1 infectivity was modest with the anti-CONE 1 sera tested, although the consistency of the effect against multiple isolates of HIV-1 Env pseudoviruses is indicative of a robust activity.
Rational protein design has reinvigorated vaccine design efforts. Scaffolded immunogens can focus immune responses to known neutralizing antibody determinants or, as in our work with the HIV-1 CONEs, can facilitate the exploitation of new vulnerable sites. The CONEs reside on an immunogenic ‘silent face’ of gp120; however, a recent study reported the identification of neutralizing antibody VRC-PG05 that recognizes several CONE 1 residues (E293, N448) as well as CONE 1 glycans (N295, N448) (Supplementary Fig. 13). This observation suggests the presence of germline antibody precursors that engage CONE-related epitopes. To achieve a functional broadly neutralizing response, a viable approach could be administering a cocktail of CONE mimetics and SOSIP Env protein missing corresponding CONE glycans in successive and combinatorial boosts, focusing the response on the ‘glycan hole’. We expect that rational design of small, thermal-stable CONE-like immunogens represents a promising starting point for the development of reproducible vaccines against persisting infections.

Methods

**Determination of the consensus sequence of subtype C gp120.** The consensus amino acid sequence of gp120 subtype C was obtained by aligning 183 full-length env sequences derived by single genome amplification. The dataset includes: 68 acute env sequences, 65 chronic env sequences, 21 functional env clone sequences from GAVI study, and 29 env sequences from viruses present in a group of participants with low CD4 counts. The amino acid position was considered a consensus if it was present in 80% of the isolates. The consensus sequences illustration for Fig. 1d was prepared using WebLogo.

We also modified the original sequence of a soluble gp140 protein (SOSIP BG505.664) so that the exposed amino acid residues of the protein that originally belonged to HIV-1 subtype A virus were replaced with residues from HIV-1 subtype C virus. The modifications were performed for the regions that correspond to CONE 1, 2, 4, and 5. Sequence manipulations were performed using BioEdit software and HIV Sequence Database Tools available at Los Alamos National Laboratory website (https://www.hiv.lanl.gov/content/sequence/HIV/mainpage.html).

**Computational protein design.** The identification of scaffold proteins for transplantation of CONE epitopes made use of the Erebus substructure search server (https://dokhlab.med.psu.edu/erebus). Query structures for this server were defined as the backbone atoms (Cα, N, Cγ) of individual residues from 12/13 (a2-helix, loop C, and loop E). We aligned the structures of Env trimer (PDB ID: 4TVP and 4NO for clade A Env, 5FYY for clade B, 5FVJ for clade G) and extracted the coordinates of backbone atoms. The query structures (PDB format) were then provided to Erebus with default search parameters (matching precision 0.5 and minimum weight ‘auto’). Erebus scans PDB for matches of structural scaffolds to atom pairs in the query. These resulting scaffolds were ranked based on their RMSD to the query structure (Supplementary Table 1).

Redesign of preexisting scaffolds was accomplished using Eris, a computational platform that automatically performs side-chain repacking and backbone relaxation and calculates the changes in free energy upon mutations (ΔΔG = ΔGmut − ΔGquery, https://dokhlab.med.psu.edu/eris). The solvent-exposed area of each residue in Env was computed with Naccess 2.1.1 (ASA value > 0.3 means exposed). The exposed residues within each CONE were introduced into the corresponding scaffolds by Eris. For each single mutation, Eris typically performed 100–1000 calculations to reach a converged distribution of ΔΔG values. The average ΔΔG and its standard deviation were obtained. Highly destabilizing mutations (ΔΔG > 6 kcal mol⁻¹) were discarded. The structures of design models were then relaxed and equilibrated by all-atom MD simulations.

To estimate the structural rigidity of CONE epitopes within designed proteins, MD simulations were performed under physiological conditions (100 ns simulation at 298 K, 150 mM NaCl, and neutral pH) with Gromacs. The force field CHARMM36 was adopted with the explicit solvent model TIP3P. The simulations were performed at constant temperature (V-rescale thermostat) and pressure (1 bar, Parrinello-Rahman NPT ensemble). The non-bonded interaction cut-off for electrostatics calculations was set to 10.0 Å, and the particle mesh Ewald (PME) method was used in the calculation of long-range electrostatic interactions. Three independent simulations were performed for each design, and the trajectories were analyzed to derive information about the average structure and RMSF.

**Protein expression and purification.** The genes encoding the designed proteins were synthesized by Biomatik and cloned into BamHI and Ncol restriction sites of pET116b vectors (Invitrogen). All constructs were confirmed by DNA sequencing.

**E. coli BL21(DE3) pLysE strains transformed with each individual gene were grown to an optical density (optical density 0.6 at 600 nm) in Lysogeny broth, and then gene expression was induced with 1 mM isopropyl-β-D-thiogalactopyranoside. Proteins expressed with an N-terminal His tag (MGHHHHHHHG) were purified on a HisTrap column (GE Healthcare) and then by size exclusion chromatography (HiPrep 16/60 Sephacryl S-200 HR). Briefly, cell pellets were resuspended in lysis buffer (20 mM NaH₂PO₄/Na₂HPO₄, 40 mM imidazole, 100 mM NaCl, 5 mM β-mercaptoethanol, pH 7.4) with protease inhibitors (phenylmethylsulfonyl fluoride and pepstatin A) and lysed by sonication. The supernatant containing protein components was separated from precipitate by centrifugation at 24,000 g for 30 min, passed through 0.22-μm filter (Millipore), and then loaded onto a HisTrap column. Proteins were eluted with a gradient (5–90%), 500 mM imidazol as 100% of imidazol in 20 mM NaH₂PO₄/Na₂HPO₄, 100 mM NaCl, pH 7.4. For the Sephacryl 200 column 30 mM NaH₂PO₄/Na₂HPO₄, 150 mM NaCl (pH 7.4) was used as running buffer. The His tag was removed by TEV protease (Sigma-Aldrich) following the manufacturer’s procedure. The cleaved proteins were enriched in flow through of the HisTrap column and were further purified through a HiLoad 16/60 Superdex 200 preparatory-grade column (GE Healthcare). The His tags were not removed from proteins conjugated to nanoparticles for immunization.

For NMR measurements, E. coli BL21(DE3) pLysE strains containing the C1S1 gene (in pET11b) were grown in M9 minimal medium with 1.0 g L⁻¹ NH₄Cl and 13C-labeled amino acids.
20 to 94 °C at 220 nm and 50,000 readings taken with 0.5-nm increment. Melting curves were monitored from before analysis. CD spectra were measured from 260 to 185 nm at 20 °C with 2.0 g L\(^{-1}\)

AnticONE 1 rabbit sera have moderate inhibitory activity against pseudotyped HIV-1. Pre- and post-immunization sera were heat inactivated, diluted 1:10 in media, and incubated with pseudotyped HIV-1 for 2 h before addition to TZM-bl cell culture. The inducible luciferase activity was quantified as a surrogate of viral infectivity. Percent neutralization (%) was determined by subtracting the pre- and post-immunization relative light units (RLU) from the virus-only RLU (not shown) and dividing the differences by the virus-only RLU. Each experiment was repeated three times. Pseudotyped HIV-1 isolates included in the assays: V682, V1086, and V3037. Rabbits ID: S1201, S1206, S1207 and 481. n = 3 independent experiments. Error bars represent standard deviation. Source data are provided as a Source Data file.

**Fig. 4** Anti-CONE 1 rabbit sera have moderate inhibitory activity against pseudotyped HIV-1. Pre- and post-immunization sera were heat inactivated, diluted 1:10 in media, and incubated with pseudotyped HIV-1 for 2 h before addition to TZM-bl cell culture. The inducible luciferase activity was quantified as a surrogate of viral infectivity. Percent neutralization (%) was determined by subtracting the pre- and post-immunization relative light units (RLU) from the virus-only RLU (not shown) and dividing the differences by the virus-only RLU. Each experiment was repeated three times. Pseudotyped HIV-1 isolates included in the assays: V682, V1086, and V3037. Rabbits ID: S1201, S1206, S1207 and 481. n = 3 independent experiments. Error bars represent standard deviation. Source data are provided as a Source Data file.

**Biophysical characterization.** CD spectroscopy data was collected using a Chirascan Plus instrument (Applied Photophysics, MA). The protein samples were dialyzed against 50 mM phosphate buffer (pH 7.4) and diluted to 0.2 mg mL\(^{-1}\) before analysis. CD spectra were measured from 260 to 185 nm at 20 °C with 50,000 readings taken with 0.5-nm increment. Melting curves were monitored from 20 to 94 °C at 220 nm and fitted to a two-state folding model for the estimation of melting temperatures.

The protein samples for size exclusion analysis (10 μM protein in 40 mM Na\(_2\)HPO\(_4\)/NaH\(_2\)PO\(_4\), 100 mM NaCl, pH 7.4) were loaded onto a Superdex 200 10/300 GL column (GE Healthcare, PA) and eluted with the same PBS buffer at a flow rate of 0.4 mL/min. UV absorption at 280 nm was monitored. Three molecular weight markers were also analyzed: cytochrome C (12.4 kDa), aprotinin (6512 Da), and vitamin B12 (1355 Da). The log of molecular weight versus elution volume for those proteins was plotted for the column and used to calculate the apparent molecular weight of designed proteins.

**Crystallization, data collection, phasing, and refinement.** Protein samples for crystallization (10–50 mg mL\(^{-1}\)) in 20 mM Tris, 150 mM NaCl, pH 8.0) were prepared by column chromatography and ultrafiltration. Sparse matrix screens in 96-well sitting drops were performed on Rigaku Phoenix Liquid Handler with MCSG Crystalization Suite. Crystallization conditions were then optimized for C2S5 and C4S3 in 24-well hanging drops. C2S5 was crystallized in 16% (w/v) polyethylene glycol 20000, 100 mM MES, pH 6.5. C4S3 was crystallized in 30% (w/v) polyethylene glycol 3350, 280 mM Bis-tris pH 6.5, and 200 mM LiSO\(_4\). Crystals were cryoprotected in the reservoir solution supplemented with 10% (v/v) glycerol, then flash frozen and stored in liquid nitrogen. The crystals were checked for high quality X-ray diffraction using Rigaku Saturn 944 + CDD with ACTOR sample.

The protein samples were concentrated by column chromatography and ultrafiltration. Sparse matrix screens in 96-well sitting drops were performed on Rigaku Phoenix Liquid Handler with MCSG Crystalization Suite. Crystallization conditions were then optimized for C2S5 and C4S3 in 24-well hanging drops. C2S5 was crystallized in 16% (w/v) polyethylene glycol 20000, 100 mM MES, pH 6.5. C4S3 was crystallized in 30% (w/v) polyethylene glycol 3350, 280 mM Bis-tris pH 6.5, and 200 mM LiSO\(_4\). Crystals were cryoprotected in the reservoir solution supplemented with 10% (v/v) glycerol, then flash frozen and stored in liquid nitrogen.
changing robot. Diffusion data was then collected on Advanced Photon Source (Argonne National Laboratory) at beamlines 22-ID and 22-BM (wavelength 0.978 Å) and processed on HKL2000.56 The structures of C255 and C455 were solved by molecular replacement with Phaser of CCP4i and PHENIX.14,42 Briefly, the crystal structures of the scaffold proteins (PDB ID 1BZ4 for C255 and 2W4C for C453) were used as search models. The structures were built and manually adjusted in Coot and then refined by Refmac5 (or by PHENIX with Composite omit map and TLS refinement options). The structures were validated using PDB validation server, Molprobity and Chiron server (https://dokhlab.med.psu.edu/chiron).44 Data collection and final refinement statistics are shown in Table 1.

NMR. For NMR measurements, 13C and 15N-enriched C1S1 protein (1 mM) was exchanged into NMR buffer (20 mM NaHPO4/Na2HPO4, pH 6.0). Five percent (v/v) D2O and 10 μM DSS (4,4-dimethyl-4-silapentane-1-sulfonic acid) were added to the protein solution. NMR spectra were acquired at 25 °C on a Bruker Avance III 850 NMR spectrometer. 2D 1H-15N HSQC experiments were recorded using 16 scans per increment and a recovery delay of 1.0 s with 2048 and 256 complex points in the direct and indirect dimensions, respectively. Spectral widths were used of 7911.393 Hz (1H) and 2152.949 Hz (15N). Average 1H-15N chemical shift perturbations were calculated according to the square root of ((Δ1H)2 + (Δ15N)2)/2, where Δ1H and Δ15N are the observed changes in 1H and 15N chemical shifts.

NMR has the potential of being a powerful tool for the diagnosis of diseases and for the study of drug targets. In this study, the authors used NMR to investigate the conformational changes in the protein of interest. The results showed that the protein underwent significant conformational changes upon binding to a particular ligand. These changes were detected by analyzing the 1H and 15N chemical shifts, which are sensitive to the local environment of the protein.

**General immunization procedure.** Immunogen injections and animal handling were performed by the Division of Laboratory Animal Medicine staff of UNC-Chapel Hill. All work with animals followed protocols that were approved by the UNC Chapel Hill Institutional Animal Care and Use Committee. Intramuscular injections were performed on 8–12-week-old New Zealand white female rabbits (2.3–4.8 kg). Two or three injections per animal were performed as immunization procedures. The details are summarized in Supplemental Table 3. The immunization scheme was an initial dose of 300 μg of each CONE followed by boost injections, supplemented with either incomplete Freund’s adjuvant according to the standard protocol, or nickel-conjugated nanoparticles without additional adjuvant. Immunogen boost injections with the same amount of protein were performed 3 weeks apart.

For CONE 1 immunization, two changes were introduced. First, the single scaffold-immunogen (C255) was compared to two scaffold-immunogen (C255 + C253) approach. In the latter approach, we alternated injections of the proteins with the same structure and HIV-specific sequence presented on different scaffolds. Second, we assessed the possibility to use the saponin-based adjuvant Matrix-M (Novavax AB) as a substitute for the nanoparticles. Matrix-M adjuvant was kindly provided by Novavax, Inc. For the first 8 weeks, 30 μg of protein were given at each injection. After the eighth dose, all animals were left in their previous condition. 10 weeks after the initial injection, the protein was administered at 300 μg of conjugated protein; therefore, the CONE 2 immunization plan was amended. After 8 weeks the amount of immunogen was increased from initial 30 μg protein to 300 μg conjugated on nanoparticles, without an adjuvant. This change of protein amount is noted in Fig. 3a.

**Immunization with nanoparticles.** Nickel nanoparticles (NPs) were prepared from warm oil/water (o/w) microemulsion precursors following a reported procedure.56,57 In the optimized formulation, Brj 71 (1.75 mg), Brj 78-NTA-Ni (1.75 mg), TPGS (1.15 mg), and Miglyol 812 (2.5 mg) were weighed into a 7 mL glass vial and heated in a water bath at 65 °C to melt and blend all excipients. A small amount of ethanol (100 μL) was added to the melted excipients and the solution was swirled to result in a homogenous mixture. The ethanol was removed completely using a stream of nitrogen and the vial was transferred to a water bath at 65 °C to form a clear, oil-in-water emulsion. The emulsion was further treated with a 1 mL of filtered and deionized water pre-heated at 65 °C, and the solution was stirred for 30 min at 65 °C then cooled to room temperature.

NPs (1 mL batches, n = 3 for each protein concentration) were characterized for particle size (NanoSight), zeta potential, and Ni content Ni content (ICP-MS) prior to injection according to the standard protocol, or nickel-free conjugated on nanoparticles, without an adjuvant. The amount of His-tagged protein (C1S1, C255, C453, or C455) at 4 °C overnight. Final protein concentration added to 0.5 mL of Ni-NPs were: 150 μg mL−1, 240 μg mL−1, 300 μg mL−1, 400 μg mL−1, and 500 μg mL−1. Free His-tagged protein was removed by spin filtration using VIVASPIN 500 ultrafiltration tubes (300 kDa MWCO). Sample containing protein and NPs was transferred to a spin filtration tube and spun at 16,000 rpm for 30 min to remove free unconjugated protein. Purified samples were analyzed by UV absorbance at 280 nm before and after spin filtration to quantify the percent conjugation of His-tagged protein to each NPs concentration (n = 3). The UV absorbance was measured using a BioTek Synergy 2 UV Spectrometer (Winooski) at a wavelength of 280 nm.

**Particle size and zeta potential.** NPs samples were run on a NanoSight NS500. All samples were diluted to a concentration between 1 × 108 and 5 × 108 particles per mL in deionized H2O. Five 60-s videos were taken of each sample to capture particles movement. The NanoSight software tracked the particles individually and using Einstein–Einstein equation, calculated the hydrodynamic diameters. The zeta potential of NPs was measure in PBS (pH 7.4) using a Malvern Zeta Sizer 2000 (Malvern Instruments).

**ICP-MS analysis.** Nickel content was quantified by inductively coupled plasma mass spectrometry (ICP-MS). The Agilent 7500cx ICPMS is outfitted with an octopole reaction cell (ORC) and a high matrix introduction (HMI) system. Standard operating conditions were Kr Power 1550 W, argon flows and plasma gas flows 15 L min−1, carrier gas flow 1.03 L min−1, make up gas 0.15 L min−1 and sampling depth was 8.0 mm. All solutions were prepared using 18 MΩ deionized water and trace metal grade nitric acid (SCP Science). The instrument was tuned daily to maximize sensitivity and minimize production of oxides and doubly charged ions. Sample flow rate was 330 μL min−1 through a Meinhard TRP-50-ANC orifice and the Scott diffusion spray chamber was cooled to 2 °C. Helium flowed at 4 mL min−1 through the ORC to eliminate isobaric interferences, 44Ca16O+ and 2Na27Cl+, of 46Ni. Standards were prepared using single element standards purchased from High Purity Standards. Scandum was used as internal standard and introduced continuously through a tee junction. Ions 65Sc (internal standard) and 60Ni were monitored in a peak hopping mode, using a 100-ms dwell time, and eight replicates were measured. The Ni standard curve included 13 concentrations levels in the range of 0.5–1000 ppb. This spanned the concentration range of all samples. The standard dataset was fitted to a linear curve. The coefficient of correlation was 0.999. Percent error in calculated concentrations was <5%. For quantitation of Ni in the NPs, preparation included removal of the water from the NPs and resuspension in 2% HNO3 solution.

**TEM imaging.** Preparations of NPs, Matrix-M adjuvant, and protein alone and in combination were negatively stained with 2% sodium phosphotungstate, pH 7.0. Five microliters of suspension was applied to a glow-discharged formvar/carbon-coated 400 mesh copper grid and allowed to adsorb for 1 min or 5 min depending on concentration. Grids were briefly floated on droplets of deionized water twice to remove buffer salts and were then transferred to a 25-μl droplet of 2% sodium phosphotungstate, pH 7.0 for 20 s. Excess stain was removed by blotting with filter paper and the grids were air dried. Grids were observed with a JEOL JEM-1230 transmission electron microscope (JEOL USA) operating at 80 kV, and digital images were acquired using a Gatan Orius SC1000 CCD camera and Gatan Microscopy Suite 3.0 software (Gatan, Inc.).
Western blot. Total rabbit IgG was purified from each animal using rProtein A GraviTrap kit (GE) according to the kit protocol, dialyzed against PBS buffer, and used at an equal amount (per CONE) for each experimental group. The gp120 pro-
teins were deglycosylated using PNGase F in SDS buffer (New England Biolabs).
Ten percent polyacrylamide gels and standard western blot protocol with SuperRock blocking reagent (ThermoFisher) was applied. The modified gp140 SOSIP-C proteins were produced at UNC Protein Expression & Puri-
cation Core. The amount of any protein loaded per lane was 400 ng. Purified IgG pools from immunized animals (1 mg ml⁻¹) were used as primary antibodies at a dilution of 1:1500. HRP goat anti-rabbit antibody (Thermo Fisher Cat #32460) was used as secondary antibody at a dilution 1:12500. The amount of any protein loaded per lane was 400 ng. Purified IgG pools from immunized animals (1 mg ml⁻¹) were used as primary antibodies at a dilution of 1:1500. HRP goat anti-rabbit antibody (Thermo Fisher Cat #32460) was used as secondary antibody at a dilution 1:12500. The fluorescence signal was detected with Western Blotting Detection Kit (GE) using a BioRad Chemidoc imager. PageRuler Plus Prestained Ladder was used for band size discrimination. To confirm the successful transfer of the gp120 proteins, the blots were stripped using Restore Western Blot Stripping Buffer (ThermoFisher).

Neutralization assays. To assess neutralizing properties of raised antibodies, we used a panel of HIV-1 pseudoviruses that expressed isolate-specific Env proteins, incubated the virus with rabbit serum, and assessed viral infectivity in the luciferase-expressing TZM-bl cells using a standard protocol (https://www.hiv. lanl.gov/content/nab-reference-strains/html/Protocol-for-Neutralizing-
Antibody-Assay-for-HIV-1-in-TZMBl_cells_Apr2017.pdf). TZM-bl cells were obtained from NIH AIDS Reagent program (cat. # 8129). Based on similarities to CONE immunogens, we chose the following HIV-1 isolates with corresponding GenBank IDs: 0665 (GB# KC984076), 0682 (GB# KC984077), 1086 (GB# KC984079), 3003 (GB# KC894087), and 3037 (GB# KC894098). In those cases where the natural isolate lacked one or more of the relatively conserved glyco-
sylation sites, these were added back to create the fully glycosylated form for that isolate. The fully glycosylated form was then mutated to remove the glycosyla-
tion site at Env positions specific for CONE proteins. For CONE 1, a glycan at the position 332 (N332S) or at the position 448 (N448S) was removed. For CONE 2, a single glycan in alpha-helix at the position 337 (N337/339S) was removed. For CONEs 4 and 5, the glycans N289S and N356/358S were removed, respectively. All positions are listed in reference to HXB2 HIV-1 viral strain. The envelope sequences of the isolates were pseudotyped on the HIV-1 subtype C viral backbone, and single-cycle infection virus was produced in HEK293T cells.

Pre- and post-immunization sera were heat inactivated and diluted 1:10 in cell

viral backbone, and single-cycle infection virus was produced in HEK293T cells.

removed. For CONEs 4 and 5, the glycans N289S and N356/358S were removed,

CONE 2, a single glycan in alpha-helix at the position 337 (N337/339S) was

removed. For CONEs 4 and 5, the glycans N289S and N356/358S were removed,

viral backbone, and single-cycle infection virus was produced in HEK293T cells.

removed. For CONEs 4 and 5, the glycans N289S and N356/358S were removed,

CONE 2, a single glycan in alpha-helix at the position 337 (N337/339S) was

removed. For CONEs 4 and 5, the glycans N289S and N356/358S were removed,

viral backbone, and single-cycle infection virus was produced in HEK293T cells.

removed. For CONEs 4 and 5, the glycans N289S and N356/358S were removed,

viral backbone, and single-cycle infection virus was produced in HEK293T cells.

removed. For CONEs 4 and 5, the glycans N289S and N356/358S were removed,

viral backbone, and single-cycle infection virus was produced in HEK293T cells.

removed. For CONEs 4 and 5, the glycans N289S and N356/358S were removed,

viral backbone, and single-cycle infection virus was produced in HEK293T cells.

removed. For CONEs 4 and 5, the glycans N289S and N356/358S were removed,

viral backbone, and single-cycle infection virus was produced in HEK293T cells.

removed. For CONEs 4 and 5, the glycans N289S and N356/358S were removed,

viral backbone, and single-cycle infection virus was produced in HEK293T cells.

removed. For CONEs 4 and 5, the glycans N289S and N356/358S were removed,

viral backbone, and single-cycle infection virus was produced in HEK293T cells.

removed. For CONEs 4 and 5, the glycans N289S and N356/358S were removed,

viral backbone, and single-cycle infection virus was produced in HEK293T cells.

removed. For CONEs 4 and 5, the glycans N289S and N356/358S were removed,

viral backbone, and single-cycle infection virus was produced in HEK293T cells.

removed. For CONEs 4 and 5, the glycans N289S and N356/358S were removed,

viral backbone, and single-cycle infection virus was produced in HEK293T cells.

removed. For CONEs 4 and 5, the glycans N289S and N356/358S were removed,
41. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D. Biol. Crystallogr.* 66, 213–221 (2010).
42. Winn, M. D. et al. Overview of the CCP4 suite and current developments. *Acta Crystallogr. D. Biol. Crystallogr.* 67, 235–242 (2011).
43. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D. Biol. Crystallogr.* 66, 12–21 (2010).
44. Jung, Y. S. & Zweckstetter, M. Mars-robust automatic backbone assignment of proteins. *J. Biomol. NMR* 30, 11–23 (2004).
45. Tamiola, K., Acar, B. & Mulder, F. A. Sequence-specific random coil chemical shifts of intrinsically disordered proteins. *J. Am. Chem. Soc.* 132, 18000–18003 (2010).
46. Benhabbour, S. R. et al. In vitro and in vivo assessment of targeting lipid-based nanoparticles to the epidermal growth factor-receptor (EGFR) using a novel Heptameric ZEGFR domain. *J. Control Release* 158, 63–71 (2012).
47. Oyewumi, M. O. & Mumper, R. J. Gadolinium-loaded nanoparticles engineered from microemulsion templates. *Drug Dev. Ind. Pharm.* 28, 317–328 (2002).

**Acknowledgements**

We thank Amy Davidson and the UNC DLAM facility for carrying out rabbit immunizations. We thank Dr. Michael Milewski, Dr. Ash Tripathy, Dr. Brenda Temple, and UNC Macromolecular Interactions Facility and UNC Protein Expression & Purification Facility staff. We thank Dr. Barton Haynes from Duke University for providing the gp120 proteins and Dr. Nancy Haigwood from Oregon Health & Science University for providing the anti-gp140 rabbit IgG solution. We thank Prof. Rogier Sanders from Weill Cornell Medical College for sharing with us the original sequence of SOSIP BG505.664 protein. We thank Ewen Liu for help with the pseudovirus panel preparation. We thank Novavax, Inc for providing us with Matrix-M adjuvant. This work was supported by NIH Grants R01 AI102732 (to N.V.D. and R.S.), GM114015 and GM123247 (to N.V.D.). Infrastructure support from the UNC Center for AIDS Research (NIH award P30 AI50410) and the UNC Lineberger Comprehensive Cancer Center (NIH award P30 CA16068) is gratefully acknowledged.

**Author contributions**

C.Z., E.D., R.S. and N.V.D. devised the project. R.S. and N.V.D. conceived the main conceptual ideas and were in charge of overall direction and planning. C.Z. designed the computational models and performed the simulations. C.Z., E.M.F., S.L.U. and J.M.F. carried out protein purification and biophysical experiments. C.Z. and H.K. obtained the crystal structures. C.Z., G.Y., and S.C. obtained and analyzed the NMR spectra. E.D., L.P., and O.C. performed characterization of rabbit antibody response to the immunogens (ELISA, IgG purification, western blots, neutralization assays). S.S.P. and S.R.B. prepared nanoparticle-conjugated immunogens. E.L.P. designed the HIV-1 pseudovirus panel. L.P.K. and E.S. assisted with rabbit sera processing and related experiments. C.Z., E.D., R.S., and N.V.D. wrote the paper with input from all authors.

**Additional information**

Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-019-08876-w.

**Competing interests:** The authors declare no competing interests.

**Reprints and permission** information is available online at http://npg.nature.com/reprintsandpermissions/

**Journal peer review information:** *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work.

**Publisher’s note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s) 2019