nature research

Corresponding author(s): P.D. Kwong & J.R. Mascola

☐ Initial submission ☒ Revised version ☐ Final submission

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size

Describe how sample size was determined.

208 HIV-1 strains were chosen to represent all major clades and circulating recombinant strains, and it is the largest panel that could be practically assessed by multiple antibodies.

58 of 208 strains (N=58) with FP sequence "AVGIGAVF" were also used to assess animal plasma and monoclonal antibody neutralization. This FP sequence is most prevalent and matches the sequence of immunogen. This panel size allowed neutralization fingerprint assessment of immune plasma and comparison with antibodies.

Initial mouse experiments were carried out with group of 3 to allow the identification of positive responses. Group of at least 5 animals were chosen for subsequent mouse, guinea pig and rhesus macaque to have 97% confidence to detect 50% positive response rate, as defined by HIV-1 neutralization.

For structural determination by cryo-EM:
(a) for the vFP1.01-DS-SOSIP complex, 1255 exposure images were collected, which after manual masking, particle picking mean standard deviation stack filtering, 2D and 3D classification, yielded 14,931 particles for final 3D reconstruction and map refinement using 3-fold symmetry.
(b) for the vFP5.01-DS-SOSIP complex, 949 exposure images were collected, which after manual masking, particle picking mean standard deviation stack filtering, 2D and 3D classification, yielded multiple 3D classes, out of which the top 3 contained 7383, 4127 and 2099 particles, and were refined with C1 symmetry.
(c) for the vFP16.02-DS-SOSIP-VRC03-PGT122 complex, 1893 exposure images were collected, which after manual masking, particle picking mean standard deviation stack filtering, 2D and 3D classification, yielded 57,278 particles for final 3D reconstruction and map refinement using 3-fold symmetry.
(d) for the vFP20.01-DS-SOSIP-VRC03-PGT122 complex, 723 exposure images were collected (using energy filter), which after manual masking, particle picking mean standard deviation stack filtering, 2D and 3D classification, yielded 48,248 particles for final 3D reconstruction and map refinement using 3-fold symmetry.

2. Data exclusions

Describe any data exclusions.

No data were excluded from the analyses in this study.

3. Replication

Describe whether the experimental findings were reliably reproduced.

All critical neutralization assays were reliably replicated as indicated in figures or methods.

All animal experiments were carried out in experimental groups. Although the magnitude varied, all of the experimental animals tested developed FP-directed antibody responses.

For vFP1-class antibodies, we describe several dozen.

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

For animal study, allocation randomization was done by weight and age.

For crystallographic refinement, reflections chosen for free-R were chosen by standard methods.

For cryo-EM experiments, "gold-standard" 3D map refinement were carried out using two models, one for each half of the data. The two half maps had their phases randomized beyond resolutions at which unmasked FSC dropped below the FSC=0.143 criterion. A mask was applied to both half maps and an FSC calculated. This FSC was used along with the original FSC before phase randomization to compute the corrected FSC as reported in Chen, S. et al. Ultramicroscopy, 135, 24-35 (2013), and implemented within RELION and cryoSparc.

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

Investigators carrying out immunizations, ELISA assays, and neutralization assessment were blinded.

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ A statement indicating how many times each experiment was replicated
☐	☒ The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☐	☒ A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒ The test results (e.g. *P* values) given as exact values whenever possible and with confidence intervals noted
☐	☒ A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒ Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.

All software are described in Methods section.

For crystal structures, diffraction data was processed with the HKL2000 suite. Structure solution was obtained by molecular replacement with Phaser. Refinement was carried out with Phenix. Model building was carried out with Coot. Structural figures were prepared with PyMOL.

For cryo-EM structures, data were acquired using Leginon. Motioncor2 was used for frame alignment and dose-weighting. CTF was estimated using the GCTF package. Particles were picked using DoG Picker within the Appion pipeline. 2D and 3D classifications were performed using RELION. After 3D classification in RELION, an additional step of ab initio reconstruction was performed using cryoSparc. 3D map refinements were carried out using either cryoSparc or RELION. Fits of HIV-1 trimer and Fab to the cryo-EM reconstructed maps were performed using Chimera. The coordinates were further fit to the electron density by an iterative process of manual fitting using Coot and real space refinement within Phenix. Molprobity and EMRinger were used to check geometry and evaluate structures at each iteration step. Figures were generated in UCSF Chimera and Pymol. Map-fitting cross

correlations were calculated using Fit-in-Map feature in UCSF Chimera. Map-to-model FSC curves were generated using EMAN2.

For MD simulation, a fully extended mannose 5 moiety at each N-linked glycosylation sequon using our in-house software glycosylator. The fusion peptide structure was then grafted onto our full mannose 5 model followed by 5000 steps of conjugate gradient energy minimization in implicit solvent using NAMD. The obtained structure was then solvated in a 17 Å padding water box, neutralized by the addition of NaCl at a concentration of 150 mM. The CHARMM force field was used for the parameterization of the protein (including CMAP corrections) and the mannose 9. TIP3P water parameterization 30 was used to describe the water molecules. Two independent molecular simulation were carried out using ACEMD molecular dynamics software on their METROCUBO workstation. Prody was used to perform principal component analysis of backbone atoms.

The associations between the sequence variability of FP neighboring glycosylation sites (HXB2 numbering 88, 241, 448 and 611) and the large panel neutralization using an in-house version of the approach implemented in R package SeqFeatR.

Prism7 was used to graph neutralization curves. Dendroscope 3 was used to visualize phylogenic tree of sequences, BioEdit v7.2.5 was used to visualize amio acid alignment, Octet software version 9.0 was used to analyze Octet binding result and generate binding curves.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.	All described materials are available by MTA for non-profit research.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).	Previously published antibodies, including VRC34.01, Motavizumab, PGT151, CH07, ACS202, VRC01, 447-52D, PGT122, VRC03, VRC01-LS, 4E10, VRC07-523-LS and VRC07-G54W, were produced in the laboratory based on published sequences. None of these antibodies have an antibody validation profile; however the antigenic profile of these antibodies, assessed by ELISA, Octet and MSD, agrees with published descriptions as referenced throughout the manuscript. Sheep D7324 antibody was purchased from AALTO Bio Reagents, with product code 7324. The lyophilised antibody was reconstituted with 2ml of double distilled water before use, as instructed by the manufacture protocol. The serological activity of the antibodies is checked by ELISA

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.	Expi293F (Invitrogen #A14527); FreeStyle™ 293-F Cells (Thermo Fisher catalog number Catalog number: R79007); HEK293S GnTI- (ATCC® CRL-3022™)
b. Describe the method of cell line authentication used.	No, we didn't authenticate the commercial cell lines.
c. Report whether the cell lines were tested for mycoplasma contamination.	Yes, mycoplasma testing was performed through Hoechst DNA staining and tested to be negative.
d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.	No commonly misidentified cell lines were used.

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.

Female mice (C57BL/6) around 8 weeks old were purchased by GenScript from Yangzhou University, CN. Female Hartley guinea pigs with body weights of 300 grams were purchased from Charles River Laboratories, MA. Female and male Indian rhesus macaques with body weights of 2-9 kg were used for immunization studies. More details were included in the methods.

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

N/A