## Supplementary information

Insights into molecular mechanism underlying CD4 dependency and neutralization sensitivity of HIV-1: A comparative molecular dynamics study on gp120s from isolates with different phenotypes

Yi Li, ${ }^{\text {ta }}$ Lei Deng, ${ }^{\text {ta }}$ Shi-Meng Ai, ${ }^{\text {b }}$ Peng Sang, ${ }^{\text {c }}$ Jing Yang, ${ }^{a}$ Yuan-Lin Xia, ${ }^{a}$ Zhi-Bi Zhang, ${ }^{a}$ Yun-Xin Fu, ${ }^{\text {a,d* }}$ Shu-Qun Liua*
a. State Key Laboratory for Conservation and Utilization of Bio-Resources in Yunnan, Yunnan University, Kunming, P.R. China
b. Department of Applied Mathematics, Yunnan Agricultural University, Kunming, P. R. China
c. College of Agriculture and Biological Science, Dali University, Dali, P. R. China
d. Human Genetics Center and Division of Biostatistics, School of Public Health, the University of Texas Health Science Center, Houston, USA
$\dagger$ These authors contributed equally to this work.

* Correspondence: Yunxin.Fu@uth.tmc.edu (YXF) shuqunliu@ynu.edu.cn (SQL)


Ramachandran Plot statistics

|  | No. of <br> residues | $\%$-tage |
| :--- | ---: | ---: |
| Most favoured regions [A,B,L] | 380 | $91.3 \%$ |
| Additional allowed regions [a,b,l,p] | 30 | $7.2 \%$ |
| Generously allowed regions [ $\sim \mathrm{a}, \sim \mathrm{b}, \sim \mathrm{\sim}, \sim \mathrm{p}]$ | 4 | $1.0 \%$ |
| Disallowed regions [XX] | 2 | $5 \%$ |
| Non-glycine and non-proline residues | 416 | $100 \%$ |

End-residues (excl. Gly and Pro) 2
Glycine residues 26
Proline residues 23

Total number of residues


Fig. S1 Ramachandran plots of the constructed structural models of H061.14- (left) and R2-gp120 (right).

Table S1. Validation results of the constructed gp120 structural models (H061.14- and R2-gp120) and template (PDB ID: 5FYJ, chain G).

| Structure or model | PROVE $^{\text {a }}$ | VERIFY3D $^{\text {b }}$ |
| :---: | :---: | :---: |
| 5FYJ, chain G | $74(5 \%)$ | $85.26 \%$ |
| H061.14-gp120 | $93(6.1 \%)$ | $79.66 \%$ |
| R2-gp120 | $101(6.5 \%)$ | $84.76 \%$ |

${ }^{\text {a }}$ The total number of buried outlier protein atoms (percentage is in parentheses).
${ }^{\mathrm{b}}$ Percent of the residues had an averaged 3D-1D score $>=0.2$.


Fig. S2. Structure-based multiple sequence alignment among the template (PDB ID: 5FYJ, chain G, labelled as 5FYJ_G), H061.14-, and R2-gp120. Conserved residues are shaded in light blue. Secondary structure (SS) is shown below the sequence alignment, with $\mathrm{H} / \mathrm{h}, \mathrm{E} / \mathrm{e}$, and $\mathrm{L} / \mathrm{I}$ representing the $\alpha$-helix (or $3 / 10$ helix), $\beta$-strand, and loop, respectively. Regular secondary structural elements of the template are numbered according to the HXBc2 crystal structures (PDB IDs: 3JWD and 1G9M), with orange arrows and red spiral representing $\beta$-strands and $\alpha$-helices (or $3 / 10$ helices), respectively. The variable regions of $\mathrm{V} 1 / \mathrm{V} 2, \mathrm{~V} 3, \mathrm{~V} 4$, and V 5 are denoted above the alignment by blue line segments. The four $\beta$-strands in the $\mathrm{V} 1 / \mathrm{V} 2$ region are labelled $\beta \mathrm{A}$ to $\beta D$, respectively, and the loops located between $\beta \mathrm{A}$ and $\beta \mathrm{B}$, between $\beta \mathrm{C}$ and $\beta \mathrm{D}$, and between $\beta \mathrm{B}$ and $\beta \mathrm{C}$ are labelled V1, V2, and L1, respectively.

