

Manuscript Number: BBAMEM-19-483

Title: CD4-binding obstacles in conformational transitions and allosteric communications of HIV gp120

Article Type: Regular Paper

Keywords: HIV envelope glycoprotein; Conformational transitions; Allosteric pathways; Molecular dynamics; Conformational selection

Corresponding Author: Dr. Yi Li,

Corresponding Author's Institution:

First Author: Yi Li

Order of Authors: Yi Li; Yu-Chen Guo; Xiao-Ling Zhang; Lei Deng; Peng Sang; Li-Quan Yang; Shu-Qun Liu

Abstract: As the only exposed viral protein at the membrane surface of HIV, envelope glycoprotein gp120 is responsible for recognizing host cells and mediating virus-cell membrane fusion. Available structures of gp120 indicate that it exhibits two distinct conformational states, called closed and open states. Although experimental data demonstrates that CD4 binding stabilizes open state of gp120, detailed structural dynamics and kinetics of gp120 during this process remain elusive. Here, two open-state gp120 simulation systems, one without any ligands (ligand-free) and the other complexed with CD4 (CD4-bound), were subjected to microsecond-scale molecular dynamics simulations. The conformational transitions and allosteric pathways of gp120 were evaluated by using the Markov state model and a network-based method, respectively. Our results provide an atomic-resolution description of gp120 conformational transitions, suggesting that gp120 is intrinsically dynamic from the open state to closed state, whereas CD4 binding blocks these transitions. Consistent with experimental structures, five metastable conformations with different orientations of the V1/V2 region and V3 loop have been extracted. The binding of CD4 significantly enhances allosteric communications from the CD4-binding site to V3 loop and β 20-21 hairpin, resulting in high-affinity interactions with coreceptors and activation of the conformational transitions switcher, respectively. This study will facilitate the structural understanding of the CD4-binding effects on conformational transitions and allosteric pathways of gp120.

Suggested Reviewers: Bing Chen
bchen@crystal.harvard.edu

Huanxiang Liu
hxliu@lzu.edu.cn

Gennady Verkhivker
verkhivk@chapman.edu

Ruth Nussinov

nussinor@mail.nih.gov



December 21, 2019

Dear editors:

We would like to resubmit our manuscript entitled "**CD4-binding obstacles in conformational transitions and allosteric communications of HIV gp120**" for your consideration of publication in *Biochimica et Biophysica Acta (BBA) - Biomembranes* as a regular paper.

This study builds on our prior paper (Li et al., 2019, *Int. J. Mol. Sci.*). Here, we applied a new method consisting of molecular dynamics simulations, Markov modeling, and network-based analysis to describe conformational transitions and allosteric communications of HIV envelope glycoprotein gp120 under two conditions, one without any ligands (ligand-free) and the other complexed with CD4 (CD4-bound). Our results reveal that gp120 is intrinsically dynamic to transfer from the open state to closed state, whereas the binding of CD4 blocks these conformational transitions. Our study will facilitate the structural understanding of the CD4-binding effects on conformational transitions and allosteric pathways of gp120.

This manuscript was submitted to BBA-Biomembranes (Manuscript No.: BBAMEM-19-314) on September 13, 2019, but unfortunately, we received a rejection decision on October 23, 2019. One of the reviewers suggests us to perform more calculations which will take more time than what is normally allowed for a major corrections decision in your journal. Considering the topic of our manuscript is within your journal scope, Dr. Hans J. Vogel encouraged us to resubmit it as a completely new submission after addressing the concerns of both reviewers in an appropriate manner (seeing more details in review_and_responses.pdf).

The first reviewer thinks our manuscript is well written, nicely illustrated, and provides a nice methodological framework for studying protein dynamics. However, she/he suggested us to extend MD simulations with glycans because gp120 is a glycoprotein. We think this issue is out the aim of our study, where two comparative MD simulation systems of open-state gp120 under the ligand-free and CD4-bound conditions were performed to investigate CD4-binding effects on the molecular dynamics, thermodynamics, and kinetics of gp120. Controlling a single variable parameter (CD4 exist or not) in the same experimental environment (open-state gp120 without glycans) can be considered as a reasonable design for our study. Nevertheless, we construct glycosylated gp120 and perform additional simulations, revealing that our results of CD4-barrier on the conformational transitions are also applicable after adding glycans. Considering the aim of current study and insufficiency of glycosylated gp120 simulations, we think it is appropriate to discuss potential effects of carbohydrates constitute in this paper and expand the MD simulations of the glycosylated gp120 as our future work.

The second reviewer approves that our study provides valuable theoretical information about conformational transitions and allosteric communications of gp120, but some specific comments should be corrected and more discussions consistent with available concepts and experimental data should be complemented. In this reversion, the conformational states of gp120 are clearly specified and more discussions about our *in silico* results consistent with experimental reports



available in numerous publications are added.

Considering the originality and innovation of this work as well as the high importance of our results to researchers in related fields, we hope that our manuscript will be reviewed again and found worthy of publication in your journal.

Thank you very much for giving us the opportunity to resubmit.

Sincerely.

A handwritten signature in black ink, appearing to be the Chinese characters '李毅' (Li Yi).

Yi Li, PhD

College of Mathematics and Computer Science
Dali University, Dali, China
Email: liyi.gerry@gmail.com

Manuscript No.: BBAMEM-19-314

Title: CD4-binding obstacles in the conformational transitions and allosteric communications of HIV gp120

Article Type: Regular Paper

Journal Title: BBA - Biomembranes

Corresponding Author: Dr. Yi Li

All Authors: Yi Li; Lei Deng; Peng Sang; Xiao-Ling Zhang; Li-Quan Yang; Shu-Qun Liu

Submit Date: Sep 13, 2019

Dear Dr. Li:

Thank you for submitting the above-named article to BBA - Biomembranes.

I am sorry to inform you that your paper is not acceptable for publication. We have completed the review of your manuscript and a summary of the comments received from two expert reviewers is appended below. Significant shortcomings were identified in the work, for example, one reviewer points out that simulations with the glycoprotein form of the protein need to be done, to provide data that are physiologically more meaningful. Addressing this issue alone would take more time than what is normally allowed for a major corrections decision in our journal. The second reviewer also has a series of concerns that would need to be addressed. Your manuscript in its present form will therefore not be considered further for publication at this stage. However, the topic is within scope and if you can address at a later date the concerns of both reviewers in an appropriate manner, you may, in the future, resubmit it as a completely new submission (in that case please include a cover letter describing the history of the manuscript).

Yours sincerely,

Hans J. Vogel, Ph.D.

Executive Editor

BBA – Biomembranes

Please see Reviewers' comments:

Reviewer #1: A manuscript by Li et al. describes simulations study of HIV protein gp120, which plays an essential role in virus entry into a host cell. Authors used a few computational methods and analysis tools, which provide a nice framework for studying protein dynamics. Thus, the manuscript is interesting from a methodological point of view. The manuscript is also well written and nicely illustrated.

There is, however, an essential problem with the biological relevance of the results. Gp120 is essentially glycoprotein; carbohydrates constitute half of the protein mass. Therefore, it is difficult to believe that glycan does not affect protein behavior. My suggestion is that authors extend simulation on glycosylated protein, including at least core glycan (5 first sugars of conserved sequence), which may represent about 30% of carbohydrates. My recommendation to reject the manuscript is based on time limit given for revision which is too short to perform such calculations.

Response:

We thank the reviewer for careful reading and constructive comments on our manuscript.

It is expected that the glycan layer could influence gp120 behavior, but this issue is beyond our research purpose. In our study, two comparative MD simulation systems of open-state gp120 under the CD4-stripped and CD4-exist conditions were performed to investigate CD4 binding effects on the molecular dynamics, thermodynamics, and kinetics of gp120. We think that controlling a single variable parameter (CD4 exist or not) in the same experimental environment (open-state gp120 without carbohydrates constitute) can be considered as a reasonable design for our study. Moreover, previous MD simulation studies on gp120 with glycosylated and non-glycosylated variable loops showed no significant differences in molecular fluctuations between these two forms of gp120 (Yokoyama et al., 2012, *PLoS One*).

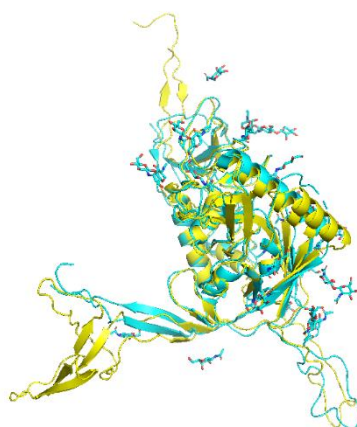


Fig. 1 A fully glycosylated gp120 (PDB ID: 6MEO, cyan) and open-state gp120 (PDB ID: 3J70, yellow). The glycans (sticks representation) stripped from 6MEO were introduced into our simulations.

However, MD simulations of the glycosylated gp120 suggested by the reviewer are interesting and worth doing. To preliminarily study how glycan affects the behavior of gp120, the glycans stripped

from a fully glycosylated gp120 (PDB ID: 6MEO, Fig. 1, cyan) (Shaik *et al.*, 2019, *Nature*) were introduced into our simulations (PDB ID: 3J70, Fig. 1, yellow). Except for using GLYCAM06 force field (Kirschner *et al.*, 2008, *J. Comput. Chem.*) to handle glycans, three 100-ns MD replicas for ligand-free and CD4-bound simulation systems were carried out under the same protocol in our manuscript.

According to time evolutions of the backbone root mean square deviation (RMSD, Fig. 2) values in MD simulations of the glycosylated gp120, ligand-free system experienced larger structural deviations and more dramatic conformational changes than the CD4-bound system, indicating that the former has a high structural flexibility and a stronger capability to alter conformation than the latter. This RMSD evaluation also implies that glycan does not influence our comparative results of the intrinsic dynamics of gp120 inhibited by the binding of CD4.

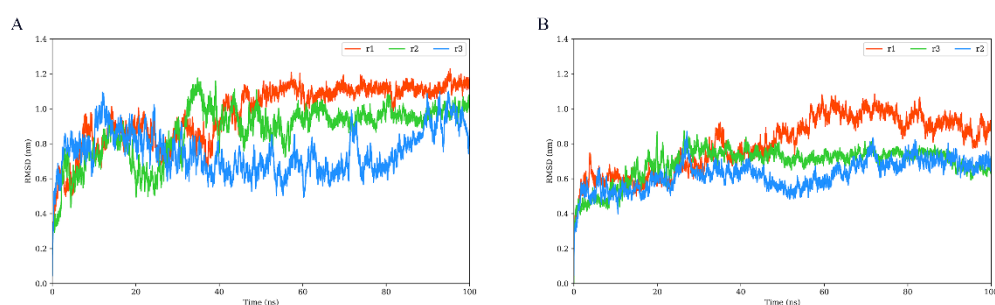


Fig. 2 Time evolution of backbone root-mean-square deviation (RMSD) values of glycosylated gp120 with respect to the starting structure calculated from three replicas (r1-3) in ligand-free (A) and CD4-bound (B) systems.

After clustering two vectors from the center of mass (COM) of the bridging sheet to the COM of the V1/V2 region and the COM of the V3 tip, a similar tendency of conformational transitions (Fig. 3) was observed in MD simulations of the glycosylated gp120. From the starting conformational state (Fig. 3A, green in ligand-free system with glycans, Fig. 3D, blue in CD4-bound system with glycans), glycosylated gp120 can sample intermediate conformation (Fig. 3C, green) only under ligand-free condition. This indicates that compared to the binding of CD4, glycans can only exert a limited effect on the conformational transitions of gp120. Our results of CD4-barrier on the conformational transitions are also applicable after adding glycans.

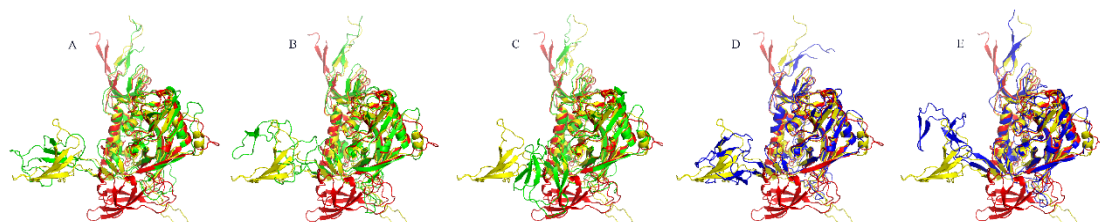


Fig. 3 Representative structures of the glycosylated gp120 in the ligand-free (A-C, blue) and CD4-bound (D-E, green) systems were superimposed to the closed-state (PDB ID: 5FYJ, red) and open-state (PDB ID: 3J70, yellow) gp120.

For the following two reasons, we think it is appropriate to add discussion about potential effects of carbohydrates constitute in this manuscript. Detailed MD simulations of glycosylated gp120 will be expanded as our future work.

1. This study we reported here focuses on the effects of CD4 binding to the molecular dynamics, thermodynamics, and kinetics of gp120. Comparative MD simulations of the ligand-free and CD4-bound gp120, although containing no glycans, could still reflect the molecular differences gp120 upon CD4 binding.
2. The MD simulations of the glycosylated gp120 we performed here are not sufficient to construct the Markov state model and can be further optimized. We look forward the reviewer to provide more professional advice for our future work.

Reviewer #2: The manuscript "CD4-binding obstacles in the conformational transitions and allosteric communications of HIV gp120" by Li et al applies molecular simulations, Markov modeling and network-based analysis to describe conformational states during the HIV-1 Env transitions from Env State 1 (closed conformation) to State 3 (open or CD4-bound conformation). The authors used computational tools and available CD4-bound Env structure as a starting point and analyzed the possible pathways by which gp120 or gp120-CD4 can transition to different states. The in silico methods/analysis provides some theoretical information but the study still needs to show consistency with concepts and experimental data available in numerous publications. Specifically, the mechanism of CD4 binding has been well investigated leading to current understanding that CD4 may bind State 1 or capture and stabilize State 3. There are also experimental data that describe intermediate states.

Response:

We thank the reviewer for careful reading and constructive comments on our manuscript.

In this revision, more descriptions and discussions consistent with the available experimental data listed by the review have been added.

Specific comments follow.

1. The term "unliganded" and "liganded" are outdated as HIV-1 gp120 is almost always crystalized with antibodies and not unliganded, and "liganded" can be ambiguous as it may represent a complex with CD4 or antibodies. The authors should use the new terms in the HIV-1 Env field that were coined in 2016 (Herschhorn et al. mBIO 2016) after the identification of a new Env intermediate state throughout the manuscript. State 1 (for the closed conformation), State 2 (intermediate conformation), and State 3 (open or CD4-bound conformation).

Response:

We now use "closed" and "open" instead of "unliganded" and "liganded", respectively, to describe the conformational states of gp120. The multiple conformational states detected in our simulations have been corresponded to the closed conformation (State 1), intermediate conformation (State 2), and open conformation (State 3) (Herschhorn et al., 2016, MBio). In order to avoid the confusion caused by the conformational states labeled by number, we use "open", "intermediate" and "closed" directly in this revision.

2. The CD4 receptor can bind to either Env conformational state. Binding to State 1 induces the transition to State 3 and binding to State 3 stabilizes this conformational state. There is not any "different role". Please delete or rephrase the sentence: "However, significant structural rearrangements between these two states and recent biophysical observations suggest that CD4 may play a different role" from the abstract. Also remove any of these concepts from the discussion section. If the authors believe that capturing State 3 is a dominant mechanism, they should provide strong experimental evidence.

Response:

We rephrase related concepts in abstract and discussion in this revision. In our study, we want to

declare capturing the open state is one of molecular mechanism. Comparative MD simulations of the ligand-free and CD4-bound systems demonstrate different conformational transitions of gp120 from open state to closed state (seeing graphical abstract, Fig. 2, and Fig. 3 in our manuscript), suggesting that gp120 is intrinsically dynamic from open state to closed state, whereas the binding of CD4 blocks these conformational transitions.

3. The authors should include the relevant references for identified Env intermediate conformations. They should at least add:

- a. Herschhorn A. et al mBIO 2016 for the intermediate state;
- b. Alsahafi N. et al Cell Host & Microbe 2019 for a new conformational state related to ADCC activity; and
- c. Lu M. et al Nature 2019 for discussion on BG505SOSIP state.

The authors should further discuss how these experimental-defined intermediates are related to their in silico Env states.

Response:

The references listed by the review have been discussed and cited in this revision.

4. The authors should validate their finding/results using available experimental data in the literature. Changes of residue 193 lead to substantial conformational changes that are consisted with the ability of L193 to form a hydrophobic core maintaining State 1. Changes to more hydrophilic residues correlate with transitions to downstream conformations. The authors should introduce in silico all amino acids to position 193, calculate the most stable conformation for each and show the correlation between opening of the trimer and hydrophobic changes. Similarly, the I423A change should stabilize State 3-like Env conformation.

Response:

Residual mutation plays a role in conformational changes of gp120, but this issue is out our research aim. In our study, two comparative MD simulation systems of open-state gp120 under the ligand-free and CD4-bound conditions were performed to investigate CD4 binding effects on the molecular dynamics, thermodynamics, and kinetics of gp120.

In fact, our group is preparing a study about molecular dynamics of gp120 from different HIV-1 isolates, which taken from an experimental report (*Seaman et al., 2010, J. Virol.*). In that study involving multiple HIV isolates will investigate the relationship between sequence variations and structural dynamics.

We have added discussion of residual mutation in this revision. We hope that the reviewers and editors will support our request to separate this issue into an independent work.

5. Delete the sentence: "It is doubtful whether such so significant structural change of gp120 between these two states should be attributed to the binding of CD4." Structural rearrangements as the result of CD4 binding are documented in numerous publications using functional, biochemical and biophysical assays.

Response:

We deleted the above sentence.

6. Rephrase "There are about 75% unliganded state and 25% liganded state at 4°C, whereas an inverted distribution of 25% unliganded state and 75% liganded state was observed at 37°C, suggesting from a thermodynamic perspective that gp120 is intrinsically able to sample a variety of conformational states." to reflect that these data is based on one method/structure. smFRET experiments show the primary HIV-1 Env are dominantly in State 1.

Response:

We deleted the description about conformational distributions at different temperatures.

7. V3 is not emanating from the bridging sheet but from under the stem of the bridging sheet.

Response:

We fixed this mistake in this reversion.

8. Please add that hydrogen-deuterium exchange analysis, which is referenced, was performed with soluble BGSOSIP trimer.

Response:

More description of hydrogen-deuterium exchange (*Guttman et al., 2014, Structure*) was added.

9. The authors used Clade G X1193.c1 SOSIP.664. They should explain why they used this structure and discuss how representative this structure is with regard to different Envs from different HIV-1 strains.

Response:

We choose the model of gp120 based on the available structure. The cryo-EM structure of HIV gp120 complexed with CD4 (PDB ID: 3J70) (*Rasheed et al., 2015, Structure*) is the only one full-length, open-state model when we carried out MD simulations. We illustrated this issue more clearly in this reversion. Other HIV strains are not considered in our study.

10. Delete or rephrase "the liganded state can be considered as a high free energy state, which can intrinsically transfer into the ground state (i.e. the unliganded state) of gp120 due to a lower free-energy level." Based on abundance of scientific data and reports, State 1 (closed conformation) occupies a HIGH energy well and CD4 binding facilitate the transition to LOWER energy state.

Response:

Evident from the single-molecule fluorescence resonance energy transfer (smFRET) experiment about conformational dynamics of gp120 (*Munro et al., 2014, Science*) shows that gp120 exhibits different conformational distributions and dominant conformations at different conditions. In the condition without any ligands (Fig. 1, unliganded), in the presence of soluble CD4 (Fig. 1, sCD4_{D1D2}),

and in the presence of sCD4_{D1D2} and 17b (Fig. 1, sCD4_{D1D2}/17b), the dominant conformation is presented by low (~ 0.10), high (~ 0.60), and intermediate (~ 0.30) FRET, respectively.

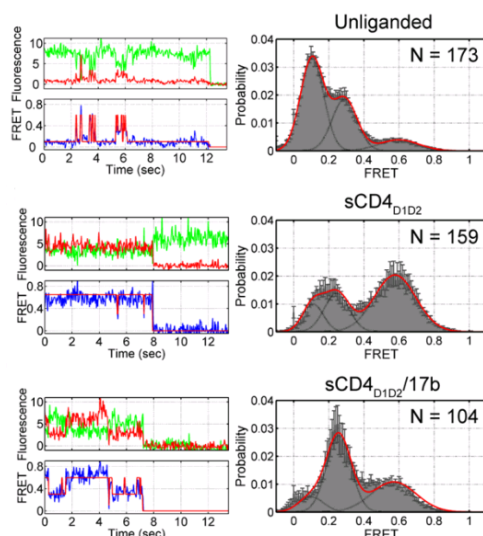


Fig. 1 Figure from single-molecule fluorescence resonance energy transfer (smFRET) experiment (Munro *et al.*, 2014, *Science*). (left) Representative fluorescence trajectories. (Right) FRET trajectories were compiled into a population FRET histogram and fit to the sum of three Gaussian distributions (red).

Based on smFRET (Fig. 1, top right), in absence any ligands, or under the “unliganded” condition, open state (high FRET) locates at a high energy level and could intrinsically transfer into the dominant, lower-energy closed state (low FRET). This supports our ligand-free simulations, which suggesting that gp120 is intrinsically dynamic from open state to closed state.

However, we rephrase the above sentence, and clarify the relationship between our simulations and the available scientific data and reports in this reversion.

11. The HR1 of gp41 is exposed on the Env surface when CD4 binds. The authors should add this information and discuss the potential changes to this region in their in silico modelling.

Response:

The possible effects of gp120 dynamics on the HR1 of gp41 have been added and discussed in this reversion.

CD4-binding obstacles in conformational transitions and allosteric communications of HIV gp120

Yi Li¹, Yu-Chen Guo¹, Xiao-Ling Zhang¹, Lei Deng², Peng Sang³, Li-Quan Yang^{3*}, and Shu-Qun Liu^{2*}

- 1. College of Mathematics and Computer Science, Dali University, Dali, China.
- 2. State Key Laboratory for Conservation and Utilization of Bio-Resources in Yunnan, Yunnan University, Kunming, China.
- 3. College of Agriculture and Biological Science, Dali University, Dali, China.

* Correspondence: shuqunliu@ynu.edu.cn (S.Q.L).

* Co-correspondence: ylbioinfo@gmail.com (L.Q.Y.).

Abstract

As the only exposed viral protein at the membrane surface of HIV, envelope glycoprotein gp120 is responsible for recognizing host cells and mediating virus-cell membrane fusion. Available structures of gp120 indicate that it exhibits two distinct conformational states, called closed and open states. Although experimental data demonstrates that CD4 binding stabilizes open state of gp120, detailed structural dynamics and kinetics of gp120 during this process remain elusive. Here, two open-state gp120 simulation systems, one without any ligands (ligand-free) and the other complexed with CD4 (CD4-bound), were subjected to microsecond-scale molecular dynamics simulations. The conformational transitions and allosteric pathways of gp120 were evaluated by using the Markov state model and a network-based method, respectively. Our results provide an atomic-resolution description of gp120 conformational transitions, suggesting that gp120 is intrinsically dynamic from the open state to closed state, whereas CD4 binding blocks these transitions. Consistent with experimental structures, five metastable conformations with different orientations of the V1/V2 region and V3 loop have been extracted. The binding of CD4 significantly enhances allosteric communications from the CD4-binding site to V3 loop and β 20-21 hairpin, resulting in high-affinity interactions with coreceptors and activation of the conformational transitions switcher, respectively. This study will facilitate the structural understanding of the CD4-binding effects on conformational transitions and allosteric pathways of gp120.

Keywords:

HIV envelope glycoprotein; Conformational transitions; Allosteric pathways; Molecular dynamics; Conformational selection.

1. Introduction

As the only exposed viral protein at the membrane surface of virion, the envelope glycoprotein gp120 of the human immunodeficiency virus (HIV) must conceal its functional centers from attack by neutralizing antibodies [1]. To strike a balance between viral infection and immune evasion, gp120 has evolved effective strategies, such as significant structural flexibility to successively bind to the receptor CD4 and co-receptors on the host cell membrane surface [2][3]. CD4 binding is the first step of a series of infective events, including the formation of co-receptor binding site, a fusion of virus-cell membranes, and the transformation of viral genetic materials [4]. With tremendous efforts over years, two distinct structures of gp120 in the closed and open states have been determined. In closed state (Fig. 1A, red), gp120 exhibits a neutralization-resistant conformation, in which the variable loops, such as the V1/V2 region and V3 loop, mask the most surface of conserved regions, and the binding sites of CD4 and coreceptors have not yet formed [5][6]. After binding to CD4, significant structural rearrangements in the open state (Fig. 1A, yellow) were observed, including dissociation of contacts between the V1/V2 region and V3 loop, re-localization of the bridging sheet, formation of CD4-induced epitopes ($\alpha 3$ and its ahead short loop), and aggregation of previously separate elements of the coreceptor binding site [7][8].

In addition to the experimental structures, the energetics and dynamics of gp120 before and after binding to CD4 were also determined. Unusual large changes in enthalpy, entropy, and heat capacity were examined by isothermal titration calorimetry (ITC) [9], detecting considerable conformational flexibility within the core of gp120. However, the properties of gp120 are not fully reflected because only a truncated form of gp120 lacking the V1/V2 region and V3 loop was used in the ITC experiment. In 2014, more detailed dynamics of gp120 upon CD4 binding were observed by the hydrogen-deuterium exchange (HDX) which measures rates of deuterium incorporation into backbone atoms in solution [10]. HDX data was used to infer the conformational flexibility of gp120 before and after adding CD4, revealing that major structural reorganizations occur in the V1/V2 region, V3 loop and bridging sheet. These two experiments imply that a series of structural rearrangements in gp120 between closed and open states are caused by the binding of CD4. However, instead of providing information during the CD4-binding process, they just compare two end-state properties before and after adding CD4.

Surprisingly, gp120 on the surface of native virions was found to be dynamic and sample at least three distinct conformations (i.e. the closed, intermediate and open states) in absence of any receptor or coreceptors by single-molecule fluorescence resonance energy transfer (smFRET) [11]. In smFRET experiment, conformational changes of gp120 were characterized by a fluorescence signal between the V1/V2 region and V4/V5 loop. Different population distributions of three gp120 conformations were observed under the different conditions (e.g. in absence of any ligands, and in presence of CD4), suggesting gp120 is intrinsically sampling multiple conformations and CD4 binding disrupts the conformational distribution of these three states of gp120 by stabilizing open state. Unfortunately, smFRET can not provide a detailed structural description of how CD4 interacts with these conformations and what effects caused by the binding of CD4 on the conformational dynamics, kinetics, and thermodynamics of gp120.

To investigate CD4-binding effects on conformational transitions and allosteric pathways of gp120, here, two open-state gp120 simulation systems, one without any ligands (ligand-free) and the other complexed with CD4 (CD4-bound), were subjected to microsecond-scale molecular dynamics (MD) simulations, in which conformational dynamics of proteins can be successfully explored [12]. Moreover, the kinetics of conformational transitions and ligand-induced allosteric regulations from extensive MD trajectories can be efficiently identified by a post-MD analysis method integrated with the stochastic process and network analysis. In this method, a Markov state model (MSM) [13] has been employed to reveal complex conformational plasticity, multistate populations, and molecular kinetics during the ligand-binding process, and a network-based approach [14] has been used to characterize allosteric regulations and pathways caused by the binding of a ligand.

In this study, open-state gp120 under ligand-free and CD4-bound conditions has been investigated by MD simulation and subsequently analyzed by the MSM and a network-based approach. It is found that gp120 is intrinsically dynamical to sample multiple conformational states and undergoes conformational transitions from the open state to closed state, whereas CD4 binding blocks these conformational transitions. The binding of CD4 not only enhances allosteric communications for the base of V3 loop, which promotes interactions with coreceptors but also shorts allosteric pathways for the β 20-21 hairpin, resulting in activation of the switcher for the conformational transitions to the open state. Our results provide detailed structural descriptions about CD4-binding effects on gp120, which will shed light on the understanding of the conformational control mechanism of HIV-1.

2. Materials and Methods

2.1. Molecular dynamics simulations

Considering the full-length gp120 structure, the starting structures of MD simulations were extracted from the cryo-EM structure of HIV gp120 complexed with CD4 (PDB ID: 3J70) [8] from Protein Data Bank (PDB) (<http://www.rcsb.org>). The ligand-free system (Fig. 1B) begins with the structure of monomeric gp120 in the open state, while the CD4-bound system (Fig. 1C) carried on a complex consisted of open-state gp120 and the D1 domain of CD4 (CD4_{D1}). To reduce computational complexity, only CD4_{D1} should be considered in the CD4-bound system due to its direct interatomic contact with gp120.

All MD simulations were performed employing GROMACS V 5.1.4 [15] with the AMBER99SB-ILDN force field [16] on GPU-accelerated clusters. Each system was solvated with TIP3P water molecules [17] and neutralized with counterions in a dodecahedron box with a protein-wall minimum distance of 1 nm. To eliminate stereochemical conflicts and soak solute into the solvent, the energy minimization with the steepest descent algorithm and the position-restrained simulation with decreasing harmonic force constants on heavy atoms of the protein were carried out. To achieve extensive conformational sampling for each system, ten of 100-ns production simulations with different initial atomic velocities assigned from the Maxwell distribution at 300 K were performed by following protocols: integration time was set as 2 fs due to LINCS [18] algorithm was used to constrain bond lengths involving hydrogen atoms; the partial-mesh Ewald (PME) [19] method and a twin-range cut-off strategy were used to calculate the long-range

electrostatic interactions and van der Waals interactions, respectively; protein and non-protein components were independently coupled to a 300 K and 1 atm with an external bath.

2.2. Markov state model

To obtain coarse-grained insights about open-state gp120 under ligand-free and CD4-bound conditions, a Markov state model (MSM) providing thermodynamic and kinetic descriptions was constructed from MD simulation trajectories by employing PyEMMA package [20]. Two appropriate features characterizing the conformational dynamics of the V1/V2 region and V3 loop were selected to discretize MD trajectories into 1447 microstates by K-means clustering. Various lag times (Fig. S1) were tested to evaluate relaxation timescales during the simulations, resulting in a lag time of 0.15 ns was chosen to construct the final MSM. Validated by the Chapman–Kolmogorov test [21] whose curves are generally in good agreements (Fig. S2), it has proven that the obtained MSM is a good approximation of these MD simulations. The perron-cluster cluster analysis (PCCA) method [22] was used to further decompose all microstates into five metastable clusters because the first gap is found after the fourth relaxation timescales. The transition probability among different microstates of gp120 was estimated by the maximum likelihood reversible transition matrix using the quadratic optimizer [23].

2.3. Network-based analysis

For each simulation system, a weighted network, where each node, edge, and weight are defined by the individual residue, the contact and cross-correlation value between pairwise residues, respectively, was constructed to describe complex molecular interactions in gp120. The C_α atom is selected to represent each residue. The contact between two residues is defined by the average distance of less than 8 Å in more than 75% of all trajectories. Each edge is weighted by cross-correlation value based on dynamic cross-correlation maps (DCCM) [14] (Fig. S3), whose elements (C_{ij}) represents the cross-correlation value between residue i and j , and can be calculated by:

$$C_{ij} = \frac{(r_i - |r_i|)(r_j - |r_j|)}{\sqrt{(r_i^2 - |r_i|^2)(r_j^2 - |r_j|^2)}}$$

where r_n is the position vectors of residue n which obtained from the aligned coordinate of the C_α atom. The value of C_{ij} can vary from -1 (completely anticorrelated motion) to +1 (completely correlated motion). To evaluate allosteric regulations between a pair of nodes in the graph of gp120, the k shortest pathways were identified using the Floyd-Warshall algorithm [24] which successfully applied in the dynamical network of biological macromolecular [25].

3. Results

3.1. Structures of gp120 and simulation systems

Numerous biophysical studies [6][7][8] provide a framework for the molecular architecture of gp120, which can be clustered into two distinct conformational states, termed as the closed and open states (fig. 1A). Except for N/C-termini, these two conformations share a highly similar core with difference in five surface-exposed variable loops (V1-5) and the bridging sheet composed of two β sheets (β2,3) and a β20-β21 hairpin. There are two important subdomains with significant structural rearrangements between the closed and open state, namely the V1/V2 region and the

V3 loop. In closed (Fig. 1A red) state, the V1/V2 region masks the V3 loop and a partial region of the CD4-binding site (CD4bs) consisted of the CD4-binding loop ($\alpha 3$ helix), \mathcal{L} D loop and the tip of the $\beta 20$ – $\beta 21$ hairpin [7]. Apparently, the V1/V2 region and the V3 loop exhibit a Y-shaped open conformation away from the core of gp120 in open state (Fig. 1B yellow), in which the coreceptor binding site, including the base and the tip of the V3 loop, has been released to help interactions with coreceptors [8]. It should be noted that the $\beta 20$ – $\beta 21$ hairpin in the bridging sheet is considered as a regulatory switcher for conformational transitions of gp120 [26].

To capture conformational transitions of gp120 from the open state to closed state, MD simulations starting with open-state gp120 extracted from gp120-ligands complex (PDB ID: 3J70 [8]) were performed. Two simulation systems of open-state gp120 without any ligands (ligand-free) and binding to CD4 (CD4-bound) were constructed to comparatively investigate conformational transitions and allosteric pathways of gp120. The ligand-free (Fig. 1B) system was started with monomeric gp120 in open state, while a simplified CD4-complexed model consisted of gp120 and the D1 domain of CD4 (CD4_{D1}) was used to construct the CD4-bound (Fig. 1C) system. It must be clear that in both simulation systems, the starting structure of gp120 is actually in the open state, regardless of the presence or absence of CD4_{D1}.

3.2. Structural deviations and sampling convergence

To assess the structural deviations of gp120 relative to the starting structure in the ligand-free and CD4-bound simulation systems, the evolution of backbone root-mean-square deviation (RMSD) values (Fig. S4) over time was calculated. In all replicas of both systems, gp120 experienced a rapid increase in RMSD values from the start of the simulation up to about until about 20 ns, eventually reaching a relatively stable equilibrium region. The equilibrium trajectory mainly distributes at 0.8 and 0.6 nm for the ligand-free and CD4-bound systems, respectively. Furthermore, the equilibrium portion of the RMSD curve clearly exhibits a narrower range for the CD4-bound gp120 than for the ligand-free gp120, with the former ranging between 0.4 and 0.9 nm and the latter between 0.5 and 1.2 nm. All above results revealed that the open-state gp120 experienced large structural deviations and more dramatic conformational changes in absence of CD4 during the simulations.

To reflect intrinsic properties of gp120, a joined equilibrium trajectory, which concatenated 20–100 ns from 10 replicas for each system was constructed due to its sufficient sampling convergence evaluated by the cosine content of the first few eigenvectors (Table S1) obtained from the principal component analysis on the MD trajectory. This value is a measure for similarity to random diffusion, ranging between 0, no cosine, and 1, a perfect cosine [27].

3.3. Conformational dynamics of the V1/V2 region and V3 loop

Since significant structural differences between the closed and open states are mainly involved in the V1/V2 region and V3 loop, the spatial orientations and conformational dynamics of these two subdomains were selected to describe the conformational transitions of gp120 during the simulations. For each snapshot of trajectories, the center of mass (COM) of the bridging sheet was used as a reference point to calculate and cluster two vectors, one to the COM of the V1/V2 region and the other to the COM of the V3 tip. The representative structure obtained from the

center of each cluster was extracted and compared to the structures of gp120 in the closed (PDB ID: 5FYJ) and open (PDB ID: 3J70) states. It should be noted that only the conformational dynamics of the V1/V2 region or V3 loop should be considered individually since this clustering is based on the description of structural orientations of the V1 /V2 region or V3 loop.

Four conformational groups were obtained (Fig. 2A) based on the clustering of vectors to the COM of the V1/V2 region, which demonstrating the conformational dynamics of the V1/V2 region relative to the core of gp120. Characterized by spatial orientations of the V1/V2 region, different significant tendencies of the gp120 conformational transitions was observed by comparing the structural differences between the representative structure from the center of each cluster and the reference structures (Fig. 2B). As the starting structure for the simulation, the open (Fig. 2B, gray) conformation of gp120 is naturally sampled as the largest cluster. In this conformation, the structural orientation of the V1/V2 region related to the core of gp120 does not change, while only different conformational arrangements occur in the V1/V2 region. Starting from the open state, gp120 in the ligand-free and CD4-bound conditions shows different tendency of conformational transitions. In the CD4-bound system, gp120 exhibits an over-open (Fig. 2B, black) conformation, in which the V1/V2 region not only moves away from the core of gp120 but also releases the most of internal regular secondary structure elements. Under the ligand-free condition, the conformational transitions of gp120 from the open state, undergoing intermediates, toward the near-closed conformational states were observed. In the structure extracted from the cluster center of intermediate (Fig. 2B, green), the spatial position of the V1/V2 region relative to the core of gp120 locates in the middle region between the closed and open states. For the near-closed (Fig. 2B, blue) conformation, the V1/V2 region has already covered the core of gp120 and is almost coincident with the closed state.

Similarly, the conformational dynamics of the V3 loop are also characterized and clustered based on spatial vectors, resulting in five structural clusters (Fig. 3A). For the sake of clarity, only the V3 loop of representative structures are displayed on the background of the starting structure. With the starting structure (Fig. 3B, yellow) as a boundary, gp120 also exhibits distinct conformational transitions under the ligand-free and CD4-bound conditions. In the ligand-free system, it can be observed that the V3 loop gradually approaches the bridging sheet from the initial open state (Fig. 3B, yellow) to the closed state (Fig. 3B, red), experiencing various intermediate states (Fig. 3B, black, green and gray). Influenced by the binding of CD4, the V3 loop can only be confined in the open state, or even transfer into the over-open state (Fig. 3B, magenta and blue).

3.4. Conformational distributions and transitions

To globally capture the conformational transitions of gp120, MD trajectories from ligand-free and CD4-bound simulation systems were combined to assemble a kinetic model of the entire process from the open state to closed state by using stochastic process. Here, an MSM estimated based on clustering of the vector characterizing the V1/2 region and V3 loop was constructed to reweigh conformational distributions and equilibrium kinetics among multiple conformations of gp120. The featurization of gp120's trajectories were discretized into 1447 microstates employing K-mean clustering and further clustered into five long-lived metastable conformations (Fig. 4).

Besides the open state (Fig. 4A), multiple conformations, such as over-open (Fig. 4B), V3-restraint (Fig. 4C), intermediate (Fig. 4D) and near-closed (Fig. 4E) conformations, can be detected. Since all simulations start from the open-state gp120, the most populated conformation, occupying 63.2% of total conformations, exhibits the near-native state (Fig. 4A), in which the representative structure of this population from the ligand-free system (Fig. 4A, blue) has a similar orientation of the V1/V2 region to open-state structural reference but the one from the CD4-bound system (Fig. 4A, green) shows a slightly over-open tendency. It should be noted that the V3 loop in this metastable conformation was released either from the representative structure of the ligand-free or CD4-bound system. Starting from this open-state conformation, gp120 exhibits different conformational transitions under different simulation conditions. In the CD4-bound system, gp120 shows two different types of conformational states. One is over-open conformation (Fig. 4B), in which the V1/V2 region moves away from the core of gp120, while the V3 loop keeps the open-like orientation. The another can be called as V3-restraint conformation, in which the V1/V2 region stabilizes in the open-state position while the V3 loop obviously was restricted in the over-open state (Fig. 4C). These two over-open conformations have a population size of 12.5% and 5.1%, respectively. In the case of the ligand-free system, a tendency from intermediate (Fig. 4D) to near-closed (Fig. 4E) state can be captured and occupy the 13.8% and 5.4%, respectively. The V3 loop keeps in the restraint position in these metastable conformations, whereas the V1/V2 region shows a movement from the open state, passing the intermediate state, to the near-closed state.

In order to obtain kinetic estimates of conformational transitions of gp120, the probability of pairwise metastable conformations was estimated. The main conformational transitions of gp120 involve from the V3-restraint conformation (Fig. 4C), the intermediate state (Fig. 4D) and the near-closed state (Fig. 4E) to the open state (Fig. 4A). The interconversion between the V3-restraint conformation (Fig. 4C) and the intermediate state (Fig. 4D) can be observed from the open state (Fig. 4A). It demonstrates that the open state is a hotspot of conformational transitions of gp120, and the direct transition from the closed state to the open state is rarely observed.

3.5. Allosteric communication

HIV evolves a two-step mechanism to enter the target cell via sequential binding of gp120 to the primary receptor CD4 and coreceptors (such as chemokine receptor CCR5). In the open state, the binding site of coreceptors, involving in proximal and distal V3 loop and partial connecting regions in the bridging sheet, are exposed [7]. The β 20-21 hairpin in the bridging sheet was considered as a major switcher of gp120 conformational transitions [26]. To evaluate CD4-binding effects to the binding of coreceptors and conformational transitions of gp120, the allosteric communication from the CD4bs to the coreceptor binding site and the switcher of conformational transitions of gp120 were analyzed by a network-based method [14], which assumes that local molecular motions lead to long-distance coupled motions in dynamics of proteins by transferring molecular energy and information between residues. In this method, a weighted network where the nodes represent residues, each edge between two nodes has been connected based on the presence ($\geq 70\%$) of contact ($\leq 8 \text{ \AA}$) between these two residues during simulations, and the weight for each edge was set as the correlation of coupled motions between residues.

Once the dynamical network has been constructed, long-distance coupled motions can be identified by searching pathways between the starting and ending points. Path lengths of 500 shortest pathways connecting a pair of residues were computed as a summation over the weights of traversed edges, in which shorter paths represent stronger molecular dynamic coupling and communications interactions. The CD4 Phe 43 pocket in the CD4bs (Glu336) of gp120 was chosen as the starting point. To identify allosteric pathways from CD4bs to the coreceptor-binding site and the switcher of conformational transitions, the center of mass (COM) of the base of V3 loop (Pro267), the tip of V3 loop (Pro279), and the β 20-21 hairpin (Ala395) were individually selected as the ending points. The allosteric pathway from CD4bs to the base of the V3 loop (Fig. 5A) and to the switcher of conformational transitions (Fig. 5C) had much shorter overall paths after binding to the CD4, whereas paths to the tip of the V3 loop (Fig. 5B) show slightly shorter in the ligand-free system. Comparing the distribution of path lengths of 500 shortest pathways in the ligand-free and CD4-bound systems indicates that the binding of CD4 significantly enhances the allosteric communication to the binding site (the V3 loop) of the coreceptor and maintains the activation of the regulatory switcher (the β 20-21 hairpin) of the conformational transitions.

4. Discussion

The *in-silico* investigation and evaluation of gp120 presented here provide an atomic description of the gp120 conformational transitions from the open state to closed state. A similar observation about conformational transitions of gp120 was reported [28], but only a truncated structure of gp120 was used. In this study, a full-length gp120 model [8] containing both the V1/V2 regions and V3 loop was employed, resulting in more detailed conformational rearrangements characterized by the structural orientations of the V1/V2 region and V3 loop. Together with our previous study about the effects of CD4 binding on the conformational dynamics, molecular motions and thermodynamics of gp120 [29], structural descriptions about CD4-binding obstacles in conformational transitions and allosteric communications have been provided to shed light on the understanding of the conformational control mechanism of HIV-1.

Our results support the observations from the smFRET experiment [11], which suggests gp120 is intrinsically dynamical among multiple conformations and the binding of CD4 reshape these conformational distributions. Based on the smFRET data, open state (high FRET) locates at a high energy level and could intrinsically transfer into the dominant, lower-energy closed state (low FRET) in absence any ligands. Similar conformational transitions from the open state to closed state were observed in our ligand-free simulations. Our study demonstrates more detailed structural descriptions about conformational transitions and provides more information about molecular dynamics, thermodynamics and kinetics during this process.

By comparing CD4-bound simulation system, it can be pointed out that CD4 binding hinders the conformation transitions of gp120 from the open to closed state, resulting gp120 has been restricted in the open state. These observations are consistent with the theory of conformational selection [30], which assumes that the native state of a protein exists a vast ensemble of conformational states, and the ligand can bind selectively to the most suitable conformational states, resulting in conformational shifting towards the particular state. Characterized by

conformational dynamics of the V1/V2 region and the V3 loop, different tendency of conformational transitions was observed under the ligand-free and CD4-bound conditions. It can be inferred that CD4 binding does not activate but block the conformational transitions of gp120 from the open state to closed state.

Besides the open, intermediate, and closed states coined in [31], multiple conformations, such as over-open, V3-restraint, and near-closed conformations have been detected from the MSM. The representative structures of these above conformations demonstrate that gp120 is structurally flexible to sample different conformational states. The V3-restraint conformation detected here exhibits similar structural characters consistency with the experimental observation in the presence of a small-molecule CD4-mimetic compound (CD4mc) [32], which can stabilize open state in combination with CD4mc. Derived from the network-based analysis, the β 20-21 hairpin, which has been considered as the regulatory switcher of the conformational transitions, receives more molecular interactions to maintain its activation after binding to CD4. It has been proposed that CD4-binding effects can be transmitted through α 1 to the HR1 of gp41, thus facilitating additional conformational changes to trigger the formation of gp41-entry machinery.

Although it can be expected that the type and size of glycans could influence the fluctuations of gp120 due to steric hindrance and weight contribution from glycans, previous MD simulation studies on gp120 with glycosylated and non-glycosylated variable loops showed no significant differences in molecular fluctuations between these two forms of gp120 [33]. In our study, two comparative MD simulation systems of open-state gp120 under the CD4-deposited and CD4-exist conditions were performed to investigate the effect of CD4 to the molecular dynamics and thermodynamics of gp120. The comparative MD simulations, although containing no glycan, could still reflect the true differences in dynamics and thermodynamics of gp120 upon CD4 binding.

5. Conclusion

In this paper, two simulation systems of open-state gp120 without any ligand (ligand-free) and complexed with CD4 (CD4-bound) were investigated by microsecond-scale molecular dynamics simulations to investigate differences in the intrinsic dynamics, conformational transitions, and allosteric communications of gp120. With the evaluation of molecular dynamics, kinetics and allosteric pathways of gp120 by the Markov state model and a network-based method, it point to a common conclusion: HIV gp120 is intrinsically dynamic to sample and transfer various conformational states, including from the open state to closed state, and the binding of CD4 hinders these conformational transitions. Characterized by the orientation of the V1/V2 region and V3 loop, ligand-free gp120 exhibits an intrinsically conformational transition from the open state to closed state, whereas CD4-bound gp120 is mainly restricted in the open state. The Markov model extracts five metastable conformations consistent with experimental structures, revealing various conformational distributions and transition probability of gp120. Based on the network-based analysis, the binding of CD4 significantly enhances the allosteric communication to the binding site (the V3 loop) of the subsequent coreceptor and maintain the activation of the regulatory switcher (the β 20-21 hairpin) of the conformational transitions. Our study about intrinsic dynamics and the role of CD4 in the viral infection will facilitate understandings of the

conformational control mechanism of HIV and may help the development of anti-HIV drugs and vaccines.

Acknowledgments

This study was funded by the Startup Foundation for Advanced Talents of Dali University (No. KYBS2018031), the Basic Research Project of Yunnan Province (No. 2019FD014 and 2019FB021), and the National Natural Sciences Foundation of China (No. 31960198).

Reference

- [1] B. Chen, Molecular Mechanism of HIV-1 Entry, *Trends Microbiol.* (2019) 1–14. doi:10.1016/j.tim.2019.06.002.
- [2] J. Liu, A. Bartesaghi, M.J. Borgnia, G. Sapiro, S. Subramaniam, Molecular Architecture of Native HIV-1 gp 120 Trimers, *Chemtracts.* 21 (2008) 227–228. doi:10.1038/nature07159.
- [3] J.B. Munro, W. Mothes, Structure and Dynamics of the Native HIV-1 Env Trimer, *J. Virol.* 89 (2015) 5752–5755. doi:10.1128/jvi.03187-14.
- [4] R. Wyatt, J. Sodroski, The HIV-1 envelope glycoproteins: Fusogens, antigens, and immunogens, *Science* (80-.). 280 (1998) 1884–1888. doi:10.1126/science.280.5371.1884.
- [5] M. Pancera, T. Zhou, A. Druz, I.S. Georgiev, C. Soto, J. Gorman, J. Huang, P. Acharya, G.Y. Chuang, G. Ofek, G.B.E. Stewart-Jones, J. Stuckey, R.T. Bailer, M.G. Joyce, M.K. Louder, N. Tumba, Y. Yang, B. Zhang, M.S. Cohen, B.F. Haynes, J.R. Mascola, L. Morris, J.B. Munro, S.C. Blanchard, W. Mothes, M. Connors, P.D. Kwong, Structure and immune recognition of trimeric pre-fusion HIV-1 Env, *Nature.* 514 (2014) 455–461. doi:10.1038/nature13808.
- [6] G.B.E. Stewart-Jones, C. Soto, T. Lemmin, G.Y. Chuang, A. Druz, R. Kong, P. V. Thomas, K. Wagh, T. Zhou, A.J. Behrens, T. Bylund, C.W. Choi, J.R. Davison, I.S. Georgiev, M.G. Joyce, Y. Do Kwon, M. Pancera, J. Taft, Y. Yang, B. Zhang, S.S. Shivatare, V.S. Shivatare, C.C.D. Lee, C.Y. Wu, C.A. Bewley, D.R. Burton, W.C. Koff, M. Connors, M. Crispin, U. Baxa, B.T. Korber, C.H. Wong, J.R. Mascola, P.D. Kwong, Trimeric HIV-1-Env Structures Define Glycan Shields from Clades A, B, and G, *Cell.* 165 (2016) 813–826. doi:10.1016/j.cell.2016.04.010.
- [7] P.D. Kwong, R. Wyatt, J. Robinson, R.W. Sweet, J. Sodroski, W.A. Hendrickson, Structure of an HIV gp 120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody, *Nature.* 393 (1998) 648–659. doi:10.1038/31405.
- [8] M. Rasheed, R. Bettadapura, C. Bajaj, Computational Refinement and Validation Protocol for Proteins with Large Variable Regions Applied to Model HIV Env Spike in CD4 and 17b Bound State, *Structure.* 23 (2015) 1138–1149. doi:10.1016/j.str.2015.03.026.
- [9] D.G. Myszka, R.W. Sweet, P. Hensley, M. Brigham-Burke, P.D. Kwong, W.A. Hendrickson, R. Wyatt, J. Sodroski, M.L. Doyle, Energetics of the HIV gp120-CD4 binding reaction, *Proc. Natl. Acad. Sci. U. S. A.* 97 (2000) 9026–9031. doi:10.1073/pnas.97.16.9026.
- [10] M. Guttman, N.K. Garcia, A. Cupo, T. Matsui, J.P. Julien, R.W. Sanders, I.A. Wilson, J.P. Moore, K.K. Lee, CD4-induced activation in a soluble HIV-1 Env trimer, *Structure.* 22 (2014) 974–984. doi:10.1016/j.str.2014.05.001.
- [11] J.B. Munro, J. Gorman, X. Ma, Z. Zhou, J. Arthos, D.R. Burton, W.C. Koff, J.R. Courter, A.B. Smith, P.D. Kwong, S.C. Blanchard, W. Mothes, Conformational dynamics of single HIV-1 envelope trimers on the surface of native virions, *Science* (80-.). 346 (2014) 759–763. doi:10.1126/science.1254426.

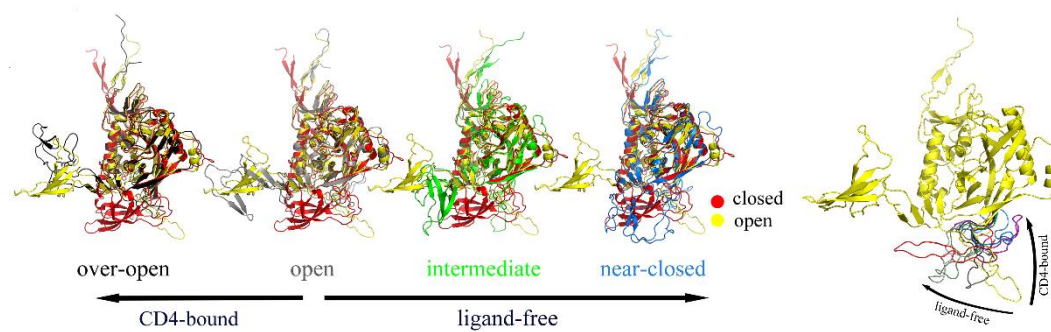
- [12] M. Karplus, J. Kuriyan, Molecular dynamics and protein function, *Proc. Natl. Acad. Sci. U. S. A.* (2005). doi:10.1073/pnas.0408930102.
- [13] B.E. Husic, V.S. Pande, Markov State Models: From an Art to a Science, *J. Am. Chem. Soc.* 140 (2018) 2386–2396. doi:10.1021/jacs.7b12191.
- [14] A. Ghosh, S. Vishveshwara, A study of communication pathways in methionyl-tRNA synthetase by molecular dynamics simulations and structure network analysis, *Proc. Natl. Acad. Sci. U. S. A.* 104 (2007) 15711–15716. doi:10.1073/pnas.0704459104.
- [15] M.J. Abraham, T. Murtola, R. Schulz, S. Páll, J.C. Smith, B. Hess, E. Lindahl, Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers, *SoftwareX.* 1–2 (2015) 19–25. doi:10.1016/j.softx.2015.06.001.
- [16] A.E. Aliev, M. Kulke, H.S. Khaneja, V. Chudasama, T.D. Sheppard, R.M. Lanigan, Motional timescale predictions by molecular dynamics simulations: Case study using proline and hydroxyproline sidechain dynamics, *Proteins Struct. Funct. Bioinforma.* 82 (2014) 195–215. doi:10.1002/prot.24350.
- [17] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, Comparison of simple potential functions for simulating liquid water, *J. Chem. Phys.* 79 (1983) 926–935. doi:10.1063/1.445869.
- [18] B. Hess, H. Bekker, H.J.C. Berendsen, J.G.E.M. Fraaije, LINCS: A Linear Constraint Solver for molecular simulations, *J. Comput. Chem.* 18 (1997) 1463–1472. doi:10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H.
- [19] T. Darden, D. York, L. Pedersen, Particle mesh Ewald: An N-log(N) method for Ewald sums in large systems, *J. Chem. Phys.* 98 (1993) 10089–10092. doi:10.1063/1.464397.
- [20] M.K. Scherer, B. Trendelkamp-Schroer, F. Paul, G. Pérez-Hernández, M. Hoffmann, N. Plattner, C. Wehmeyer, J.H. Prinz, F. Noé, PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models, *J. Chem. Theory Comput.* 11 (2015) 5525–5542. doi:10.1021/acs.jctc.5b00743.
- [21] F. Noé, C. Schütte, E. Vanden-Eijnden, L. Reich, T.R. Weikl, Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations, *Proc. Natl. Acad. Sci. U. S. A.* 106 (2009) 19011–19016. doi:10.1073/pnas.0905466106.
- [22] P. Deuffhard, M. Weber, Robust Perron cluster analysis in conformation dynamics, *Linear Algebra Appl.* 398 (2005) 161–184. doi:10.1016/j.laa.2004.10.026.
- [23] J.H. Prinz, H. Wu, M. Sarich, B. Keller, M. Senne, M. Held, J.D. Chodera, C. Schütte, F. Noé, Markov models of molecular kinetics: Generation and validation, *J. Chem. Phys.* 134 (2011). doi:10.1063/1.3565032.
- [24] R.W. Floyd, Algorithm 97: Shortest path, *Commun. ACM.* 5 (1962) 345. doi:10.1145/367766.368168.
- [25] A. Sethi, J. Eargle, A.A. Black, Z. Luthey-Schulten, Dynamical networks in tRNA: Protein complexes, *Proc. Natl. Acad. Sci. U. S. A.* 106 (2009) 6620–6625. doi:10.1073/pnas.0810961106.
- [26] A. Herschhorn, C. Gu, F. Moraca, X. Ma, M. Farrell, A.B. Smith, M. Pancera, P.D. Kwong, A. Schön, E. Freire, C. Abrams, S.C. Blanchard, W. Mothes, J.G. Sodroski, The β 20- β 21 of gp120 is a regulatory switch for HIV-1 Env conformational transitions, *Nat. Commun.* 8 (2017). doi:10.1038/s41467-017-01119-w.
- [27] B. Hess, Convergence of sampling in protein simulations, *Phys. Rev. E - Stat. Physics, Plasmas,*

- Fluids, Relat. Interdiscip. Top. 65 (2002). doi:10.1103/PhysRevE.65.031910.
- [28] A. Korkut, W.A. Hendrickson, Structural Plasticity and Conformational Transitions of HIV Envelope Glycoprotein gp120, PLoS One. 7 (2012). doi:10.1371/journal.pone.0052170.
- [29] Y. Li, L. Deng, L.Q. Yang, P. Sang, S.Q. Liu, Effects of CD4 binding on conformational dynamics, molecular motions, and thermodynamics of HIV-1 gp120, Int. J. Mol. Sci. 20 (2019) 260. doi:10.3390/ijms20020260.
- [30] X. Du, Y. Li, Y.L. Xia, S.M. Ai, J. Liang, P. Sang, X.L. Ji, S.Q. Liu, Insights into protein–ligand interactions: Mechanisms, models, and methods, Int. J. Mol. Sci. 17 (2016) 144. doi:10.3390/ijms17020144.
- [31] A. Herschhorn, X. Ma, C. Gu, J.D. Ventura, L. Castillo-Menendez, B. Melillo, D.S. Terry, A.B. Smith, S.C. Blanchard, J.B. Munro, W. Mothes, A. Finzi, J. Sodroski, Release of GP120 restraints leads to an entry-competent intermediate state of the HIV-1 envelope glycoproteins, MBio. (2016). doi:10.1128/mBio.01598-16.
- [32] N. Alsahafi, N. Bakouche, M. Kazemi, J. Richard, S. Ding, S. Bhattacharyya, D. Das, S.P. Anand, J. Prévost, W.D. Tolbert, H. Lu, H. Medjahed, G. Gendron-Lepage, G.G. Ortega Delgado, S. Kirk, B. Melillo, W. Mothes, J. Sodroski, A.B. Smith, D.E. Kaufmann, X. Wu, M. Pazgier, I. Rouiller, A. Finzi, J.B. Munro, An Asymmetric Opening of HIV-1 Envelope Mediates Antibody-Dependent Cellular Cytotoxicity, Cell Host Microbe. (2019). doi:10.1016/j.chom.2019.03.002.
- [33] M. Yokoyama, S. Naganawa, K. Yoshimura, S. Matsushita, H. Sato, Structural dynamics of HIV-1 envelope GP120 outer domain with V3 loop, PLoS One. (2012). doi:10.1371/journal.pone.0037530.

Highlights

- HIV gp120 is intrinsically dynamic to sample various conformational states.
- An atomic-resolution description of gp120 conformational transitions from the open to closed state was provided by microsecond-scale molecular dynamics simulations.
- The binding of CD4 hinders the gp120 conformational transitions.
- CD4-binding obstacles on the conformational transitions of gp120 can be obtained from the allosteric communications.

Graphical Abstract



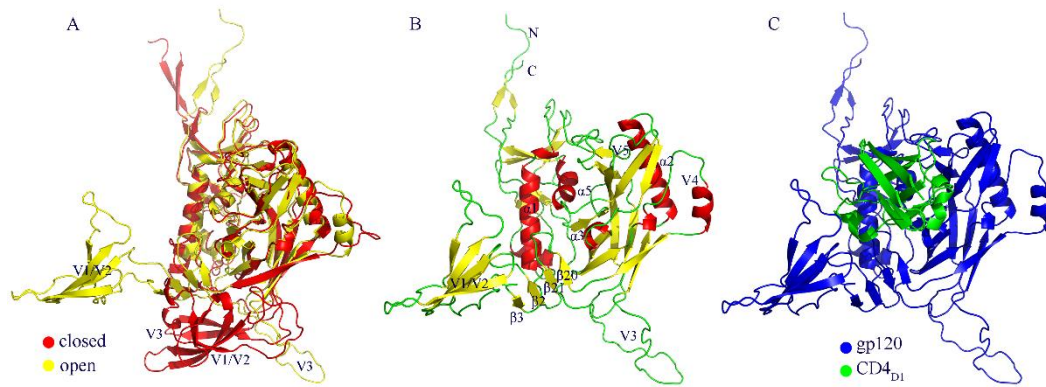


Fig. 1 Structural models of gp120 and the starting structures of simulation systems. (A) Structural superimpose of gp120 in the closed (PDB ID: 5FYJ, red) and open (PDB ID: 3J70, yellow) states. (B) Ribbon representation of open-state gp120 (PDB ID: 3J70, chain D) is used as the starting structure for the ligand-free simulation system. (C) The starting structural model of the CD4-bound simulation system consists of open-state gp120 (blue) and the D1 domain of CD4 (CD4_{D1}, green).

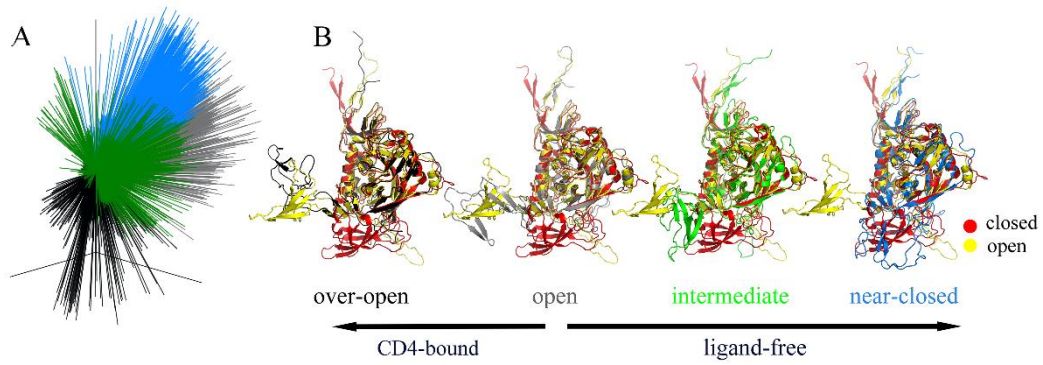


Fig. 2 Orientations of the V1/2 region. (A) Clustering on vectors calculated from the center of mass (COM) of the bridging sheet to the COM of the V1/V2 region for every snapshot of trajectories. (B) Four representative structures selected from the center of each cluster were labeled as over-open (black), open (gray), intermediate (green) and near-closed (blue). Each representative structure was superimposed to the closed-state (PDB ID: 5FYJ, red) and open-state (PDB ID: 3J70, yellow) gp120.

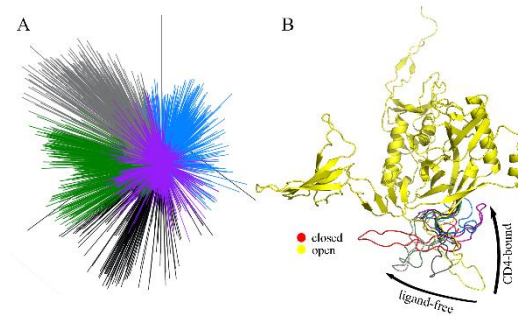


Fig. 3 Conformations of the V3 loop. (A) Clustering on vectors calculated from the center of mass (COM) of bridging sheet to the COM of the V3 loop for every snapshot of trajectories. (B) Five representative structures (only the V3 loop were shown) from the center of each cluster were superimposed to the structure of closed-state (PDB ID: 5FYJ, red, only the V3 loop were shown) and open-state (PDB ID: 3J70, yellow) gp120.

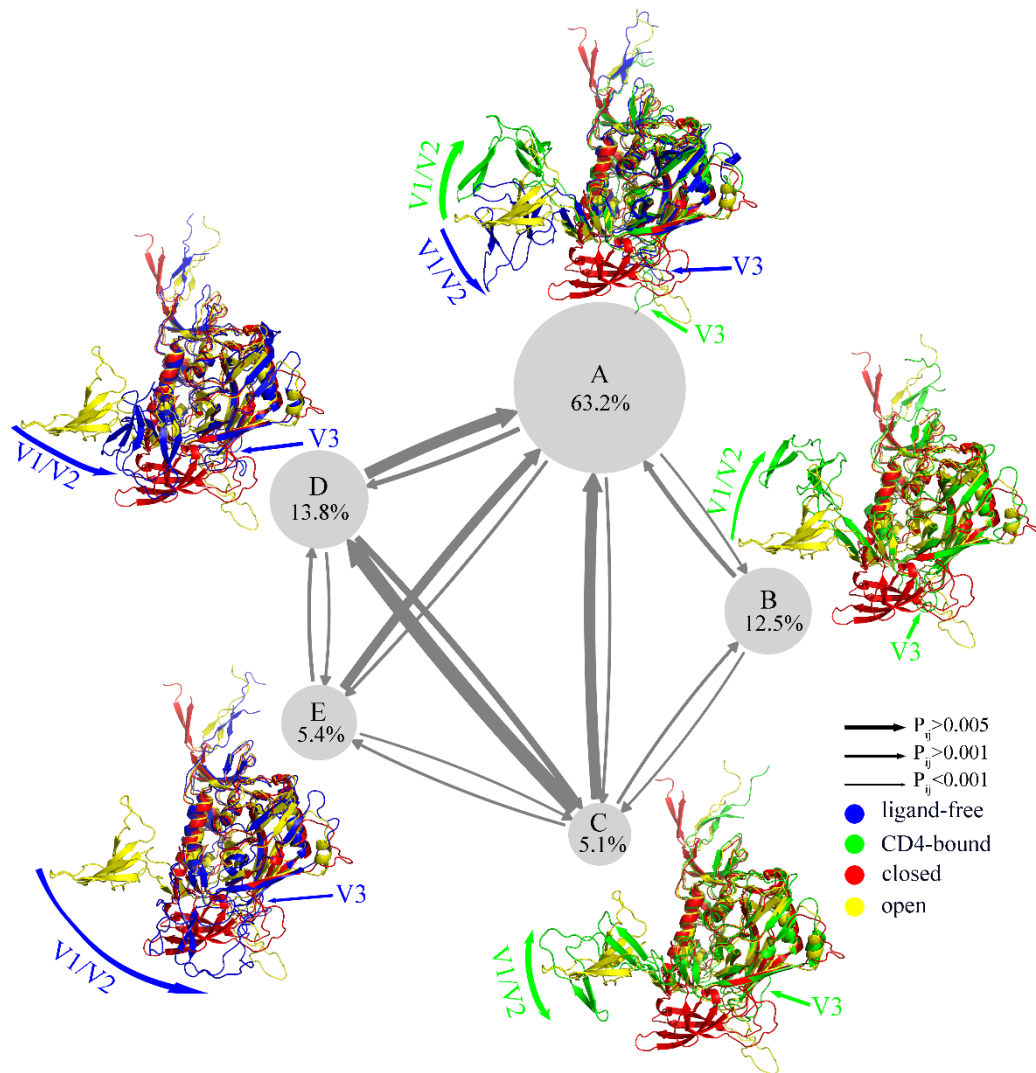


Fig. 4 Distributions and kinetics of five metastable conformations extracted from the Markov state model. Representative structures selected from the center of each metastable conformation from the ligand-free (blue) and CD4-bound (green) simulation systems were superimposed to closed-state (PDB ID: 5FYJ, red) and open-state (PDB ID: 3J70, yellow) gp120. Each circle has an area proportional to the distribution of conformational population, which has been marked in each circle. Every arrow with a different width indicates the probability of a conformational transition.

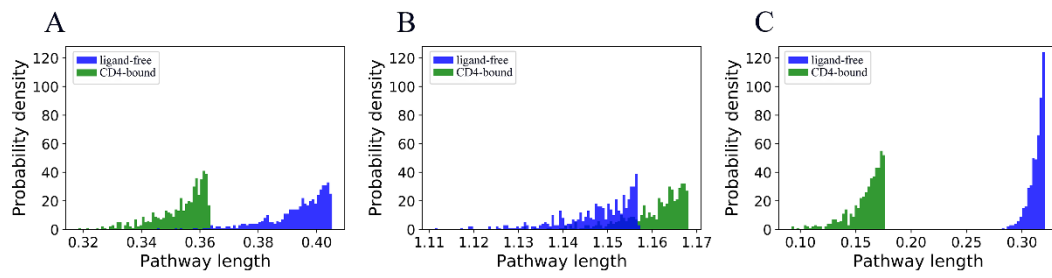


Fig. 5 Length of allosteric pathways. In the ligand-free (blue) and CD4-bound (green) simulation systems, the probability density of length of the 500 shortest allosteric pathways from the CD4-binding site to the base (A) and tip (B) of the V3 loop, and the β 20-21 hairpin (C) was calculated.

CD4-binding obstacles in conformational transitions and allosteric communications of HIV gp120

Supplementary data

Yi Li¹, Yu-Chen Guo¹, Xiao-Ling Zhang¹, Lei Deng², Peng Sang³, Li-Quan Yang^{3*}, and Shu-Qun Liu^{2*}

1. College of Mathematics and Computer Science, Dali University, Dali, China.

2. State Key Laboratory for Conservation and Utilization of Bio-Resources in Yunnan, Yunnan University, Kunming, China.

3. College of Agriculture and Biological Science, Dali University, Dali, China.

* Correspondence: shuqunliu@ynu.edu.cn (S.Q.L).

* Co-correspondence: ylbioinfo@gmail.com (L.Q.Y.).

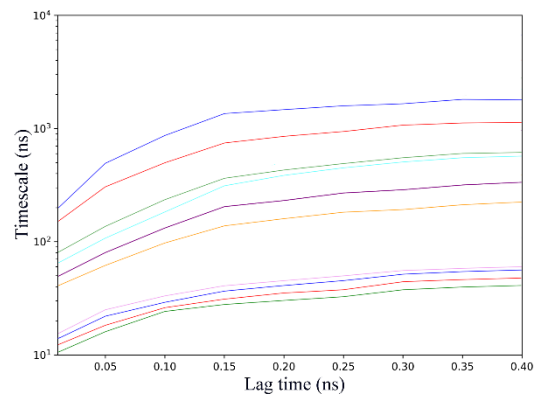


Fig. S1 Relaxation timescales. Different lag time were test to estimate the relaxation timescale.

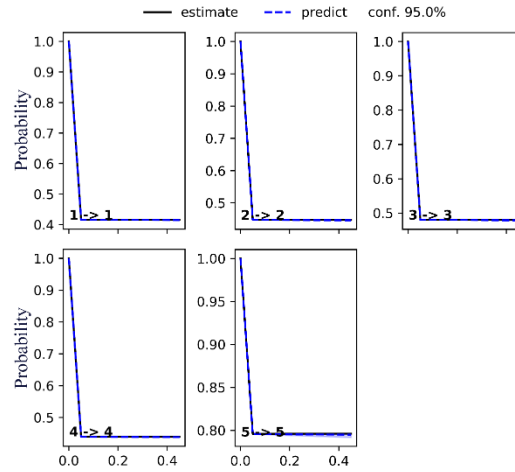


Fig. S2 A Chapman-Kolmogorov test of the Markov state model. Each graph shows the evolution of the population of a given metastable state (1-5, corresponding to A to E in the main text Fig. 4).

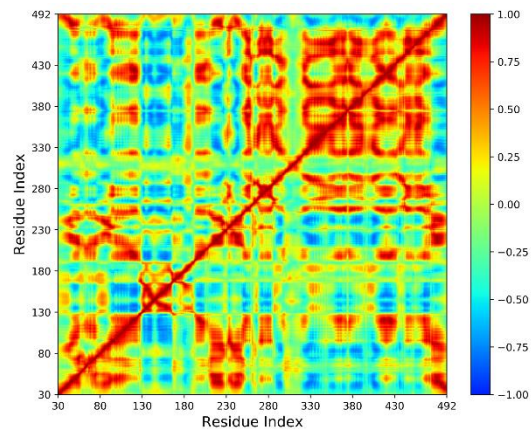


Fig. S3 Dynamic cross-correlation map (DCCM). The correlation coefficient values of C_{α} motions of the ligand-free and CD4-bound gp120 during simulations showed in the upper and lower triangle of DCCM, respectively.

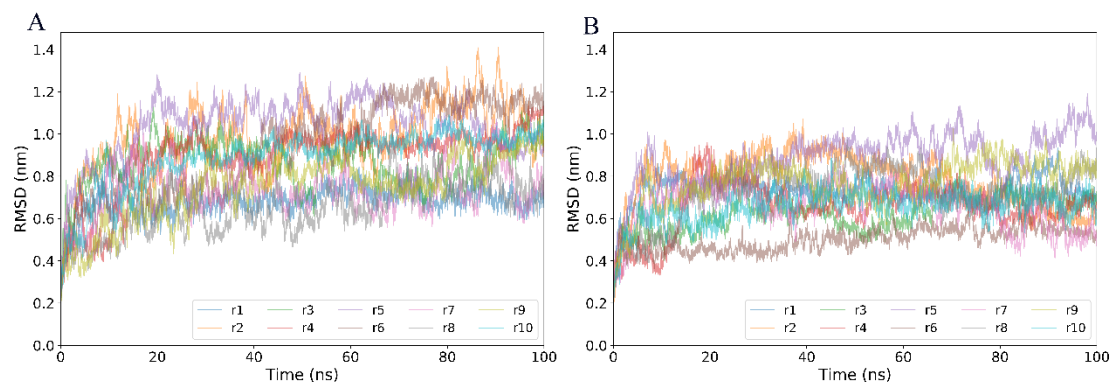


Fig. S4 Molecular dynamics of gp120. Time evolution of backbone root-mean-square deviation (RMSD) values of gp120 with respect to the starting structure calculated from ten replicas (r1-10) in ligand-free (A) and CD4-bound (B) systems.

Table S1 Cosine content values of the first three eigenvectors calculated from replicas and joined equilibrium trajectories.

	ligand-free system			CD4-bound system		
	Eigenvector 1	Eigenvector 2	Eigenvector 3	Eigenvector 1	Eigenvector 2	Eigenvector 3
Replica 1	0.0390	0.0796	0.0518	0.6624	0.2412	0.3600
Replica 2	0.2944	0.0954	0.2729	0.0234	0.2040	0.0327
Replica 3	0.6562	0.3313	0.0495	0.5688	0.0642	0.1059
Replica 4	0.8122	0.7335	0.6163	0.0506	0.6243	0.3437
Replica 5	0.4338	0.0868	0.1488	0.4934	0.0016	0.0807
Replica 6	0.6938	0.0505	0.0041	0.5208	0.0222	0.0107
Replica 7	0.1069	0.0981	0.0018	0.3998	0.3055	0.0016
Replica 8	0.7286	0.3991	0.0264	0.6430	0.6253	0.4043
Replica 9	0.3194	0.0215	0.0068	0.5999	0.3918	0.1189
Replica 10	0.7268	0.3253	0.0431	0.7029	0.2496	0.0298
Joined	0.0164	0.0044	0.0093	0.0137	0.0748	0.0070

Please wait...

If this message is not eventually replaced by the proper contents of the document, your PDF viewer may not be able to display this type of document.

You can upgrade to the latest version of Adobe Reader for Windows®, Mac, or Linux® by visiting http://www.adobe.com/go/reader_download.

For more assistance with Adobe Reader visit <http://www.adobe.com/go/acrreader>.

Windows is either a registered trademark or a trademark of Microsoft Corporation in the United States and/or other countries. Mac is a trademark of Apple Inc., registered in the United States and other countries. Linux is the registered trademark of Linus Torvalds in the U.S. and other countries.