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

Yi Li[1], Lei Deng[2], Peng Sang[3], Xiao-Ling Zhang[1], 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: ylqbioinfo@gmail.com (L.Q.Y.).

Abstract

The envelope glycoprotein gp120, the only exposed viral protein at the membrane surface of HIV, is responsible for recognizing host cells and mediating virus-cell membrane fusion. Available structures of gp120 indicate that it exhibits two distinct conformational states, termed as the unliganded and liganded states. It is natural to attribute the conformational changes of gp120 to the binding of a ligand (e.g. CD4). However, significant structural rearrangements between these two states and recent biophysical observations suggest that CD4 may play a different role. Here, two simulation systems of liganded-state gp120, one is without any ligands (ligand-free) and the other binds to CD4 (CD4-bound), were subjected to microsecond-scale molecular dynamics simulations following the evaluation of conformational transitions and allosteric pathways of gp120 by using the Markov state model and a network-based method, respectively. Our results provide an atomic-resolution description of gp120 conformational transitions, demonstrating that gp120 is intrinsically dynamic from the liganded state to the unliganded state, whereas the binding of CD4 blocks these conformational transitions. In this process, five metastable conformations with different orientations of the V1/V2 region and the V3 loop have been extracted. The binding of CD4 significantly enhances allosteric communications from the CD4 binding site to the V3 loop and the β20-21 hairpin, resulting in high-affinity interactions with coreceptors and the activation of the conformational transitions switch, 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 the virion, the envelope glycoprotein gp120 of the human immunodeficiency virus (HIV) must maintain its functional centers to recognize host cells while conceal them from attack by neutralizing antibodies [1]. To strike a balance between viral infection and immune evasion, HIV evolves some effective strategies, such as significant structural flexibility to successively bind to receptor CD4 and coreceptors on the surface of host cells [2][3]. The binding of CD4 is the first step of a series of infective events, including the formation of the coreceptor binding site, a fusion of virus-cell membranes, and the transformation of genetic materials [4]. With tremendous efforts over twenty years, two distinct structures of gp120 in the unliganded and liganded states have been determined [5][6][7][8][9]. In the unliganded state (Fig. 1A, red), gp120 exhibits a neutralization-resistant conformation, in which the glycan-shield variable loops (the V1/V2 region and the V3 loop) mask most of the exposed regions, and the binding sites of CD4 and coreceptors have not yet formed [6][7]. Despite sharing a similar core region, significant structural rearrangements in the liganded state (Fig. 1A, yellow) were observed, including dissociation of contacts between the V1/V2 region and the V3 loop, re-localization of the bridging sheet, formation of CD4-induced epitopes (α3 and its ahead short loop), and aggregation of previously separate elements of the coreceptor binding site [5][10].

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) [11], revealing considerable conformational flexibility within the core of gp120. However, the properties of gp120 are not fully reflected because only a deglycosylated and truncated form of gp120 lacking the V1/V2 region and the V3 loop was used in the ITC experiment. In 2014, more detailed dynamics of gp120 upon CD4 binding was observed by the hydrogen-deuterium exchange (HDX) which measures the rates of deuterium incorporation into backbone amides in solution [12]. The HDX data was used to comparatively infer the conformational flexibility of gp120 before and after adding CD4, revealing that major structural reorganizations occur in the V1/V2 region, the V3 loop and the bridging sheet. These two experiments seem to indicate that a series of structural rearrangements in

gp120 between the unliganded and liganded 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. It is doubtful whether such so significant structural change of gp120 between these two states should be attributed to the binding of CD4.

Surprisingly, gp120 on the surface of native virions was found to be dynamic and sample at least three distinct conformations (i.e. the unliganded, intermediate and liganded states) in absence of any receptor or coreceptors by single-molecule fluorescence resonance energy transfer (smFRET) [13]. In the smFRET experiment, the conformational changes of gp120 were characterized by fluorescence signal between the V1/V2 region and the V4/V5 loop. Different population distributions of these three conformations were observed under the different conditions, suggesting the binding of CD4 disrupts the conformational distribution of these three states of gp120 and promotes gp120 transfer out the unliganded state into the liganded state. Unfortunately, smFRET does 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, molecular motions, and thermodynamics of gp120. Another evident showing dynamics of gp120 comes from recent cryogenic electron microscopy (cryo-EM) study, which found that the sampling temperature influences the conformational distribution of gp120 [8]. There are about 75% unliganded state and 25% liganded state at 4℃, whereas an inverted distribution of 25% unliganded state and 75% liganded state was observed at 37℃, suggesting from a thermodynamic perspective that gp120 is intrinsically able to sample a variety of conformational states.

However, the above experiments cannot provide a clear atomic-level picture for conformational transitions of gp120. The molecular mechanism underlying conformational transitions between these two states and CD4-binding effects during this process still remains elusive. One possible way to investigate conformational flexibility of gp120 and the CD4-binding role is to investigate conformational flexibility of liganded-state gp120 without any ligand (ligand-free) and binding to CD4 (CD4-bound) by employing the molecular dynamics (MD) simulation, in which conformational dynamics of proteins can be successfully explored. Moreover, efficiently

identifying processes and kinetics of conformational transitions and ligand-induced allosteric regulations from extensive MD trajectories can be obtained by a post-MD analysis method integrated with the stochastic process and network analysis. In this method, a Markov state model (MSM) [14] has been employed to reveal complex conformational plasticity, multistate population, and kinetics during the ligand-binding process, and a network-based approach [15] has been used to characterize allosteric regulations and pathways caused by the binding of the ligand.

In this study, the liganded-state gp120 under the 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 liganded state to the unliganded state which is blocked by the binding of CD4. The binding of CD4 not only enhances allosteric communications for the base of the V3 loop which promotes interactions with coreceptors but also shorts allosteric pathways for the β20-21 hairpin, resulting in the activation of the switch for the conformational transitions to the liganded 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

The starting structures of MD simulations were extracted from the cryo-EM structure of HIV gp120 complexed with CD4 (PDB ID: 3J70) [10]. The ligand-free system begins with the structure of monomeric gp120 in the liganded state, while the CD4-bound system carried on a complex consisted of liganded-state gp120 and D1 domain of CD4 ($CD4_{D1}$). To reduce computational complexity, the only $CD4_{D1}$ should be considered in the CD4-bound system due to its direct interatomic contact with gp120. All MD simulations were performed by employing GROMACS [16] V5.1.4 with the AMBER99SB-ILDN [17] force field on GPU-accelerated clusters. Each system was solvated with TIP3P [18] water molecules 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. For each system, ten replicas of 100-ns production simulations starting from random initial atomic velocities assigned from the Maxwell distribution at 300 K were performed following protocols: integration time was set as 2 fs due to LINCS [19] algorithm was used to constrain bond lengths involving hydrogen atoms; the partial-mesh Ewald (PME) [20] 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 of liganded-state gp120 under the ligand-free and CD4-bound conditions, a Markov state model (MSM) providing thermodynamic and kinetic descriptions was constructed from MD simulation trajectories by combining functions of the PyEMMA software [21]. Two appropriate features characterizing the conformational dynamics of the V1/V2 region and the V3 loop (more detailed in Results) were selected to discrete MD trajectories into 1447 microstates by K-means clustering. Various lag time was tested (Fig. S1) to evaluate relaxation timescales during simulations, resulting in a lag time of 0.15 ns was chosen to construct the final MSM. By validated by the Chapman–Kolmogorow test [22] whose curves are generally in good agreement (Fig. S2), the Markov model is a good approximation of the underlying simulations. The perron cluster cluster analysis (PCCA) method [23] was used to further decompose all microstates into five metastable clusters according to a gap is found after fourth relaxation timescales. The transition probability among different microstates of gp120 was estimated by the maximum likelihood reversible transition matrix using the quadratic optimizer [24].

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 the 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 is defined by between two average distance

less than 8 Å and the proportion of all trajectories larger than 75%. Each edge is weighted by cross-correlation value based on dynamic cross-correlation maps (DCCM) [15] (Fig. S3), whose elements (C_{ij}) represents the cross-correlation value between residue i and j, and can be calculated by:

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

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 [25] which successfully applied in the dynamical network of biological macromolecular [26]. Comparative path length distributions indicating the strength of correlated motions under distinct conditions were then calculated.

3. Results

3.1. Structures of gp120 and simulation systems

Numerous biophysical studies [7][5][10][8] provide a framework for the molecular architecture of gp120, which can be clustered into two distinct conformational states, termed as the unliganded and liganded states (fig. 1A). Except for N/C-termini, these two conformations share a highly similar core with the difference in the five surface-exposed variable loops (V1-5) and the bridging sheet composed of the β2-3 and β20-β21 hairpins. Emanated from the bridging sheet, there are two important subdomains, namely the V1/V2 region and the V3 loop. In unliganded (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 (α3 helix), $\mathcal{L}$D loop and the tip of the β20-β21 hairpin [5]. Apparently, the V1/V2 region and the V3 loop exhibit a Y-shaped open conformation away from the core of gp120 in the liganded 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 [10]. It is should be noted that the β20–β21 hairpin in the bridging sheet is considered as a regulatory switch for conformational transitions of gp120 [27] undergoes a roll from the order of β2-β3-β21-β20 in the unliganded state to the β3-β2-β21-β20 in liganded states.

To capture conformational transitions of gp120 from the liganded state to the unliganded state, MD simulations starting with liganded-state gp120 extracted from gp120-ligands complex (PDB ID: 3J70 [10]) were performed. Two simulation systems of liganded-state gp120 without any ligands and binding to CD4 were constructed to comparatively investigate conformational flexibility and allosteric pathways of gp120. The ligand-free (Fig. 1B) system was started with monomeric gp120 in the liganded 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 liganded 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, mainly distributed 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 of RMSD values 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 liganded-state gp120 experienced large structural deviations and more dramatic conformational changes in the absence of CD4 during the simulation.

To ensure 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. It has been shown that insufficient sampling should be considered when the cosine content of the first few

eigenvectors is close to 1 [28].

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

Since significant structural differences between the unliganded and the liganded states are mainly involved in the V1/V2 region and the V3 loop, the spatial orientations and conformational dynamics of these two subdomains of gp120 were selected to describe the conformational transitions of gp120 during the simulation. For each snapshot of trajectories, the center of mass (COM) of the bridging sheet was used a reference point to calculate and cluster two vectors one from to the COM of the V1/V2 region and the other from 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 unliganded (PDB ID: 5FYJ) and liganded (PDB ID: 3J70) states. It should be noted that only the conformational dynamics of the V1/V2 region or the V3 loop should be considered individually since this clustering is based on the description of structural orientations of the V1 /V2 region or the V3 loop.

Based on the clustering of the conformational dynamics of the V1/V2 region relative to the core of gp120, four conformational groups were obtained (Fig. 2A). 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 liganded (Fig. 2B, gray) conformation of gp120 is naturally sampled as the largest clustering population. 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 liganded state, gp120 in the ligand-free and CD4-bound conditions shows different conformational transitions. In the CD4-bound system, gp120 exhibits an over-ligand (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. For the ligand-free system, the conformational transitions of gp120 from the liganded state, undergoing intermediates, toward the near-unliganded conformational states were observed based on the spatial orientations and conformational

dynamics of the V1/V2 region. 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 at the middle region between the unliganded and liganded states. For the near-unliganded (Fig. 2B, blue) conformation, the V1/V2 region has already covered the core of gp120 and is almost coincident with the unliganded state.

Similarly, the conformational dynamics of the V3 loop are also characterized and clustered based on spatial vectors, obtaining five structural clusters (Fig. 3A). For the sake of clarity, only the V3 loop of the reference structure and the representative structures is displayed on the background of the starting structure since it is always covered by the V1/V2 region. With the starting structure 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 liganded state (Fig. 3B, yellow) to the unliganded 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 to the liganded state, or even transfer into the over-liganded 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 over-liganded state to the near-unliganded state by using appropriate 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 liganded state (Fig. 4A), over-liganded (Fig. 4B), V3-restraint (Fig. 4C), intermediate (Fig. 4D) and near-unliganded (Fig. 4F) conformations can be detected. Since all simulations start from the liganded-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 liganded-state structural reference but the one from the CD4-bound system (Fig. 4A, green) shows a slightly over-liganded tendency. It should be noted that the V3 loop in this metastable conformation was released either from the representative structure of ligand-free or CD4-bound system. Starting from this liganded-state conformation, gp120 exhibits different conformational transitions under different simulation condition. In the CD4-bound system, gp120 show two different types of conformational states. One is over-liganded conformation (Fig. 4B), in which the V1/V2 region move away from the core of gp120, while the V3 loop keeps the liganded-like orientation. The another can be called as V3-restraint conformation, in which the V1/V2 region stabilizes in the liganded-state position while the V3 loop obviously was restricted in the over-liganded state (Fig. 4C). These two over-liganded 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-unliganded (Fig. 4E) state can be captured and occupy the 13.8% and 5.4%, respectively. The V3 loop keeps in restraint position in these metastable conformations, whereas the V1/V2 region shows a movement from the liganded state, passing the intermediate state, to the near-unliganded 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-unliganded state (Fig. 4E) to the liganded state (Fig. 4A). The interconversion between the V3-restraint conformation (Fig. 4C) and the intermediate state (Fig. 4D) can be observed from the liganded state (Fig. 4A). It demonstrates that the liganded state is a hotpot of conformational transitions of gp120, and the direct transition from the unliganded state to the liganded 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 a coreceptor (such as chemokine receptor CCR5). In the liganded state,

the binding site of coreceptors, involving in proximal and distal V3 loop and partial connecting regions in the bridging sheet, are exposed [5]. Otherwise, the β20-21 hairpin in the bridging sheet was considered as a major swicher of gp120 conformational transitions [27]. 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 [15], which assumes that local coupled 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 (≥ 70%) of contact (≤ 8 Å) 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 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 β20-21 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 maintain the activation of the regulatory switcher (the β20-21 hairpin) of the conformational transitions.

4. Discussion and conclusion

The *in silico* investigation and evaluation of gp120 presented here provide an atomic description of the gp120 conformational transitions from the liganded state to the unliganded state. A similar observation about conformational transitions of gp120 was reported [29], but only the truncated structure of gp120 was used. In this study, a full-length gp120 model [10] containing both the V1/V2 regions and the V3 loop was employed, resulting in more detailed conformational rearrangements characterized by the structural orientations of the V1/V2 region and the V3 loop were extracted during this process. Together with our previous study about the effects of CD4 binding on the conformational dynamics, molecular motions and thermodynamics of gp120 [30], 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.

By comparative analysis of molecular dynamics and conformational transitions of gp120 under the ligand-free and CD4-bound conditions, it can be inferred that the binding of CD4 does not activate but block the conformational transitions of gp120 from the liganded state to the unliganded state. These results support the observations from smFRET [13] which suggests gp120 is intrinsically dynamical among multiple conformations and the binding of CD4 reshape these conformational distributions. Our studies provide an explanation for how the binding of CD4 changes the conformational populations of gp120. The liganded-free simulation system has demonstrated that gp120 can intrinsically transfer out high-free-energy liganded state into the ground unliganded state. By comparing CD4-complexed simulation system, it can be pointed out that the binding of CD4 hinders the conformation transition of gp120 from liganded to unliganded state. It is noteworthy that the binding of CD4 can change the conformational distributions of gp120, but it is doubtful that CD4 induces unliganded-state gp120 into the liganded state. Our study, together with the smFRET data, suggests that the binding of CD4 puts obstacles in the gp120 conformational transitions from the liganded to the unliganded states, resulting gp120 has been restricted in the liganded state. These observations are consistent with the theory of conformational selection [31], 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.

In this paper, two simulation systems of liganded-state gp120 without any ligand (ligand-free) and binding to CD4 (CD4-bound) were investigated by microsecond-scale molecular dynamics simulations to investigate the 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 liganded state to the unliganded state, and the binding of CD4 hinders these conformational transitions. Characterized by the orientation of the V1/V2 region and the V3 loop, ligand-free gp120 exhibits an intrinsically conformational transition from the liganded state to the unliganded state, whereas CD4-bound gp120 is mainly restricted in the liganded state. The Markov model extracts five metastable conformations with different structural features, 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 while inhibiting the regulatory switcher (the β20-21 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.31860243 and 31660015), and the National Natural Sciences Foundation of China (NO. 31960198, 31860243, and 31660015).

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 gp120 trimers, Nature. (2008). doi:10.1038/nature07159.

[3] J.B. Munro, W. Mothes, Structure and Dynamics of the Native HIV-1 Env Trimer, J. Virol. (2015). doi:10.1128/jvi.03187-14.

[4] R. Wyatt, J. Sodroski, The HIV-1 envelope glycoproteins: Fusogens, antigens, and immunogens, Science (80-.). (1998). doi:10.1126/science.280.5371.1884.

[5] 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.

[6] 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. (2014). doi:10.1038/nature13808.

[7] 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. (2016). doi:10.1016/j.cell.2016.04.010.

[8] G. Ozorowski, J. Pallesen, N. De Val, D. Lyumkis, C.A. Cottrell, J.L. Torres, J. Copps, R.L. Stanfield, A. Cupo, P. Pugach, J.P. Moore, I.A. Wilson, A.B. Ward, Open and closed structures reveal allostery and pliability in the HIV-1 envelope spike, Nature. (2017). doi:10.1038/nature23010.

[9] M.M. Shaik, H. Peng, J. Lu, S. Rits-Volloch, C. Xu, M. Liao, B. Chen, Structural basis of coreceptor recognition by HIV-1 envelope spike, Nature. 565 (2019) 318–323. doi:10.1038/s41586-018-0804-9.

[10] 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. (2015). doi:10.1016/j.str.2015.03.026.

[11] 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. 97 (2002) 9026–9031. doi:10.1073/pnas.97.16.9026.

[12] 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. (2014). doi:10.1016/j.str.2014.05.001.

[13] 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-.). (2014). doi:10.1126/science.1254426.

[14] B.E. Husic, V.S. Pande, Markov State Models: From an Art to a Science, J. Am. Chem. Soc. (2018). doi:10.1021/jacs.7b12191.

[15] 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. (2007). doi:10.1073/pnas.0704459104.

[16] M.J. Abraham, T. Murtola, R. Schulz, S. Páll, J.C. Smith, B. Hess, E. Lindah, Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers, SoftwareX. (2015). doi:10.1016/j.softx.2015.06.001.

[17] 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. (2014). doi:10.1002/prot.24350.

[18] 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. (1983). doi:10.1063/1.445869.

[19] B. Hess, H. Bekker, H.J.C. Berendsen, J.G.E.M. Fraaije, LINCS: A Linear Constraint Solver for molecular simulations, J. Comput. Chem. (1997). doi:10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H.

[20] T. Darden, D. York, L. Pedersen, Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems, J. Chem. Phys. (1993). doi:10.1063/1.464397.

[21] 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. (2015).

doi:10.1021/acs.jctc.5b00743.

[22] 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. (2009). doi:10.1073/pnas.0905466106.

[23] P. Deuflhard, M. Weber, Robust Perron cluster analysis in conformation dynamics, Linear Algebra Appl. (2005). doi:10.1016/j.laa.2004.10.026.

[24] J.H. Prinz, H. Wu, M. Sarich, B. Keller, M. Senne, M. Held, J.D. Chodera, C. Schtte, F. Noé, Markov models of molecular kinetics: Generation and validation, J. Chem. Phys. (2011). doi:10.1063/1.3565032.

[25] R.W. Floyd, Algorithm 97: Shortest path, Commun. ACM. (1962). doi:10.1145/367766.368168.

[26] A. Sethi, J. Eargle, A.A. Black, Z. Luthey-Schulten, Dynamical networks in tRNA:protein complexes, Proc. Natl. Acad. Sci. (2009). doi:10.1073/pnas.0810961106.

[27] 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 β20-β21 of gp120 is a regulatory switch for HIV-1 Env conformational transitions, Nat. Commun. (2017). doi:10.1038/s41467-017-01119-w.

[28] B. Hess, Convergence of sampling in protein simulations, Phys. Rev. E - Stat. Physics, Plasmas, Fluids, Relat. Interdiscip. Top. (2002). doi:10.1103/PhysRevE.65.031910.

[29] A. Korkut, W.A. Hendrickson, Structural Plasticity and Conformational Transitions of HIV Envelope Glycoprotein gp120, PLoS One. (2012). doi:10.1371/journal.pone.0052170.

[30] 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.

[31] 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.

Highlights

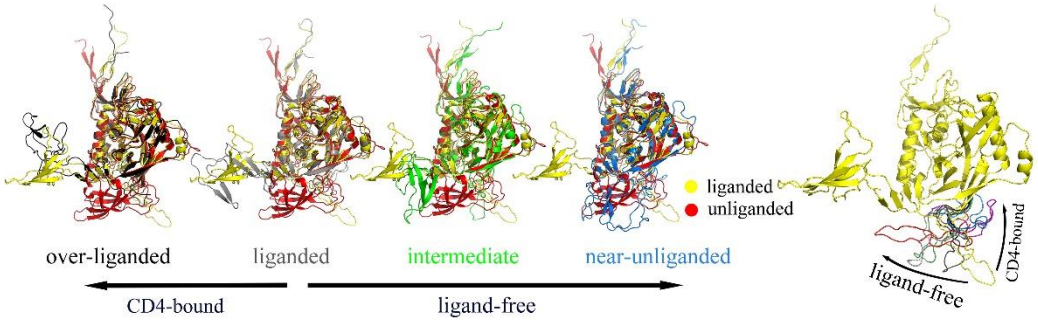
• HIV gp120 is intrinsically dynamic to sample various conformational states.

• An atomic-resolution description of gp120 conformational transitions from the liganded to the unliganded state.

• 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



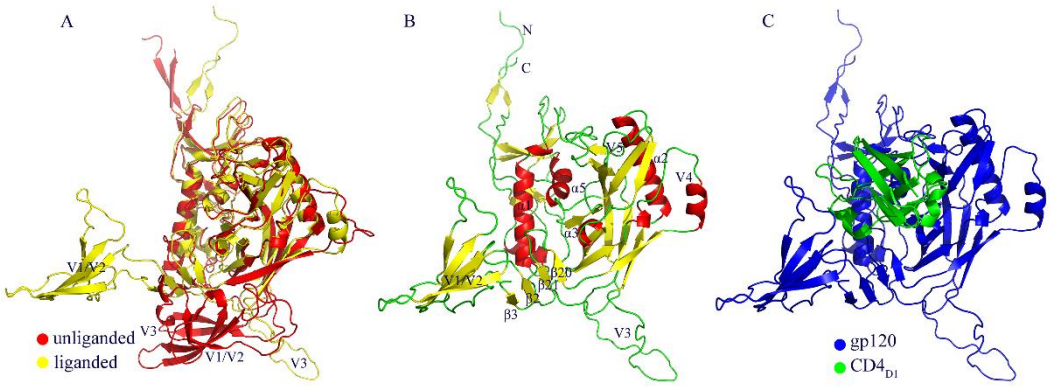


Fig. 1 Structural models of gp120 and simulation systems. (A) Ribbon representation of liganded gp120 (PDB ID: 3J70) used as the starting structure of the ligand-free simulation systems. (B) Structural superimpose of gp120 in the unliganded (PDB ID: 5FYJ, red) and liganded (yellow) states. (C) The starting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

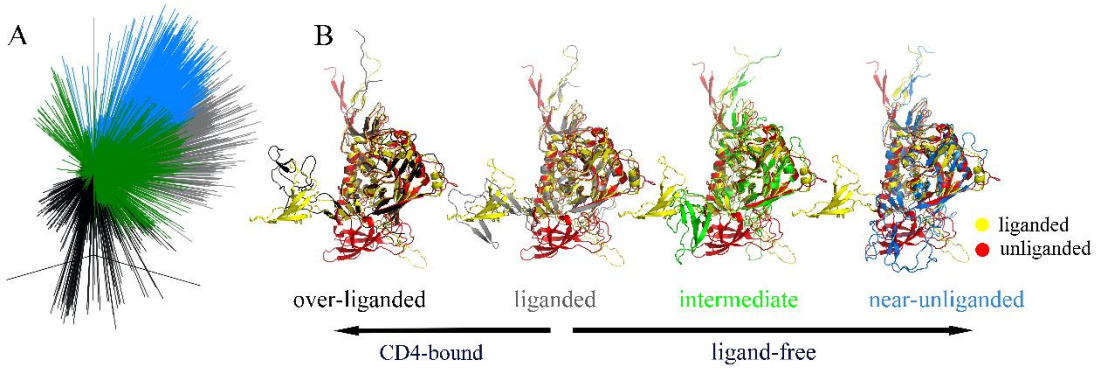


Fig. 2 Orientations of the V1/2 region. (A) Clustering on vectors calculated from the center of mass (COM) of the bridging sheet to the COM of the V1/V2 region for every snapshot of trajectories. (B) Four representative structures were selected from the center of each cluster and labeled as over-liganded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

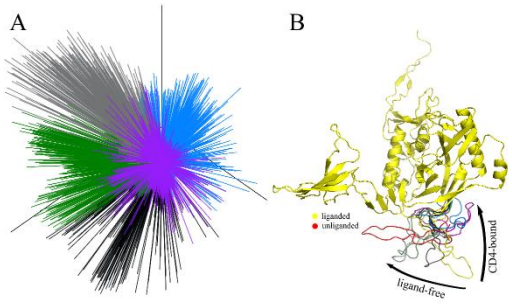


Fig. 3 Conformations of the V3 loop. (A) Clustering on vectors calculated from the center of mass (COM) of bridging sheet to the COM of the V3 loop for every snapshot of trajectories. (B) Five representative structures (only the V3 loop were shown) from the center of each cluster were superimposed to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

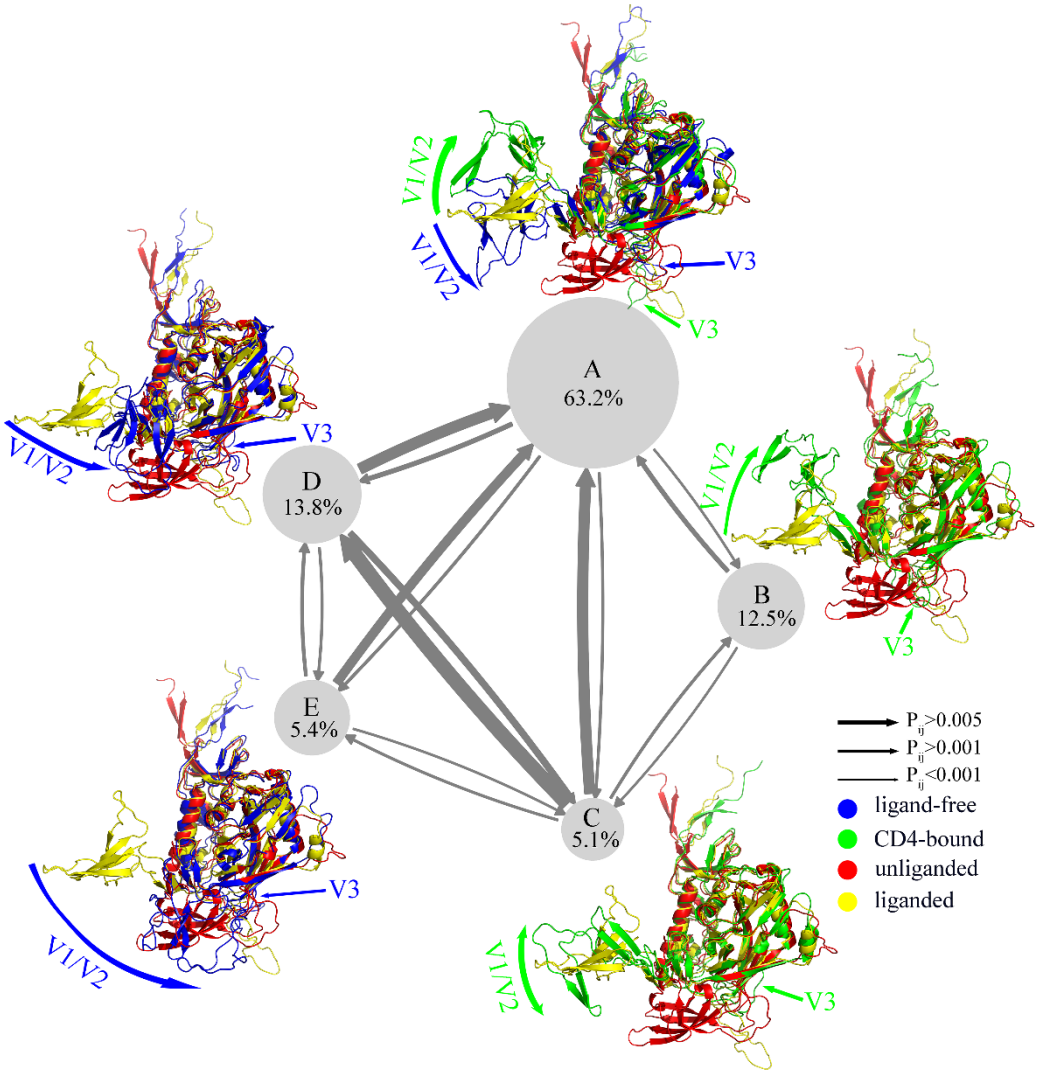


Fig. 4 Distributions and kinetics of five metastable conformations extracted from the Markov state model. Each circle has an area proportional to the distribution of conformational population which has been marked in the respective circle. Each arrow with different width indicates the probability of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

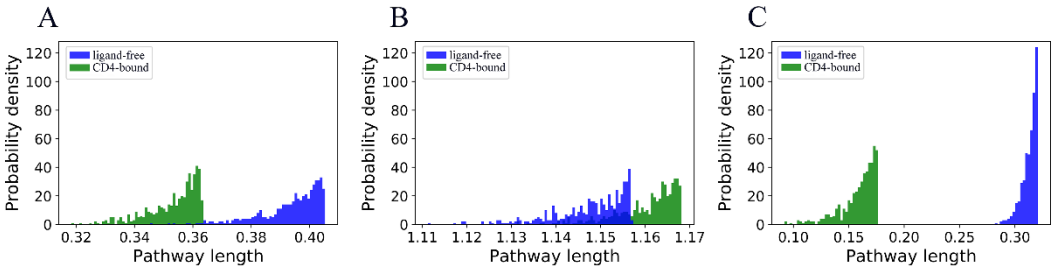


Fig. 5 The length of allosteric pathways. For the ligand-free (blue) and CD4-bound (green) gp120, the probability density of the length of the 500 shortest allosteric pathways from CD4-binding site to the base (A) and tip (B) of the V3 loop, and the β20-21 hairpin (C) was calculated.