J Mol Model (2008) 14:857–870
DOI 10.1007/s00894-008-0327-7

ORIGINAL PAPER

Molecular motions of human HIV-1 gp120 envelope glycoproteins

Shu-Qun Liu · Shi-Xi Liu · Yun-Xin Fu

Received: 17 February 2008 / Accepted: 29 May 2008 / Published online: 2 July 2008

Abstract The HIV-1 gp120 exterior envelope glycoprotein undergoes a series of conformational rearrangements while sequentially interacting with the receptor CD4 and the coreceptor CCR5 or CXCR4 on the surface of host cells to initiate virus entry. Both the crystal structures of the HIV-1 gp120 core bound by CD4 and antigen 17b, and the SIV gp120 core pre-bound by CD4 are known. Despite the wealth of knowledge on these static snapshots of molecular conformations, the details of molecular motions crucial to intervention remain elusive. We presented a comprehensive comparative analysis of dynamic behavior of gp120 in its CD4-complexed, CD4-free and CD4-unliganded states based on the homology models with modeled V3 and V4 loops. CONCOORD computer simulation was utilized to generate ensembles of feasible protein structures, which were subsequently analyzed by essential dynamics technique to identify preferred concerted motions. The revealed collective fluctuations are dominated by complex motional modes such as rotation/twisting, flexing/closing, and shortness/elongation between or within the inner, outer, and bridging-sheet domains. An attempt has been made to relate these modes to receptor/coreceptor association and neutralization avoidance. Covariance web analysis revealed four subdomains that undergo concerted motion in gp120. The structural components in gp120 that move in concert with CD4 were also identified, which may be the suitable target for inhibitor design to interrupt CD4-gp120 interaction. The differences in B-factors between the three gp120 states revealed certain structural regions that could be related either to CD4 association or to subsequent dissociation of gp120 from gp41. These dynamics data provide new insights into the structure-function relationship of gp120 and may aid in structure-based anti-HIV vaccine design.

Keywords B-factor · Collective fluctuation · Comparative modeling · CONCOORD simulation · Covariance web · Essential dynamics · HIV-1 gp120

Abbreviations

HIV	Human immunodeficiency virus
AIDS	Acquired immunodeficiency syndrome
Fab	Antigen binding fragment
SIV	Simian immunodeficiency virus
V	Variable loop
MD	Molecular dynamics
ED	Essential Dynamics
PANC	Per-atom normalized covariance
MSF	Mean square fluctuation
RMSD	Root mean square deviation
CD4-BL	CD4-binding-loop

S.-Q. Liu · Y.-X. Fu (✉)
Laboratory for Conservation and Utilization of Bio-resources,
Yunnan University,
Kunming 650091, People's Republic of China
e-mail: shuqunliu@ynu.edu.cn

S.-X. Liu
School of Chemical Science and Technology, Yunnan University,
Kunming 650091, People's Republic of China

Y.-X. Fu
Human Genetics Center,
The University of Texas Helth Science Center,
Houston, TX 77030, USA

Introduction

Over 40 million people are currently infected with the human immunodeficiency virus type 1 (HIV-1), the major cause of the acquired immune deficiency syndrome (AIDS) [1, 2]. The development of a preventive vaccine, which optimally should elicit both virus-neutralizing antibodies and cellular immune response, is of high priority and urgency [3, 4].

Entry of HIV-1 into target cells requires transformation of the protective envelope into a fusion-competent state. Infection is initiated by the selective interaction between the viral exterior envelope glycoprotein, gp120, and receptor on the target cell, CD4, and obligatory chemokine receptors (CCR5 or CXCR4) [5–8]. The envelope glycoprotein precursor gp160 polypeptide chain is cleaved during transport process into the gp120 and gp41, which are organized into trimetric complexes on the viron surface [9, 10]. Accumulating biochemical and structural evidences indicate that the initial binding of CD4 to gp120 triggers conformational changes in the HIV envelope, which subsequently promote recognition of coreceptors and ultimately lead to conformational changes of gp41 and membrane fusion [8, 10]. This complex process involves a series of structural rearrangements in which conformational dynamics of the HIV envelope glycoprotein plays a crucial role.

Recently, the crystallographic structure of the gp120 core in complex with two amino-terminal domain (D1D2) of CD4 and the antigen binding fragment (Fab) of the human neutralizing antibody, 17b, has been determined in a series of X-ray crystallographic studies [11, 12]. It turns out that the gp120 core is composed of the inner and outer domains and the bridging sheet. Components of both domains and bridging sheet contribute to CD4 binding; and several structural elements of this core appear to depend on association with CD4 for their stabilization. Although the crystal structure of the CD4-bound HIV-1 gp120 core is incomplete due to residue deletions from both the N- and C- termini; Gly-Ala-Gly tripeptide substitutions for 67 V1/V2 loop residues and 32 V3 loop residues; and the lack of electronic density map for the V4 loop, this core provides tremendous structural insights for analyzing envelope antigenicity and infection mechanism [11–13]. Another advance is a more recent determination of the crystallographic structure of the SIVmac 32H gp120 in its CD4-unliganded state. This structure reveals gp120's prefusion state, before interaction with CD4 [14]. The unliganded SIV gp120 core is also composed of the inner and outer domains and contains neither the variable loops V1/V2 nor the V3. However, the V4 loop was resolved, which adopts an open conformation, extending away from the body of outer domain. A rough comparison between the structures of the unliganded SIV gp120 and the CD4-bound HIV gp120 indicates that conformational changes involve dramatic reconfiguration and translocation of structural components in the inner domain and bridging sheet, but not in the outer domain [14]. The unexpectedly extensive conformational rearrangement upon receptor binding is consistence with the findings from thermodynamics studies, suggesting that the binding of CD4 (or antibodies) to gp120 induces substantial structural rearrangements primarily in the core structure [15, 16].

Although the crystallographic structures provide important snapshots of the gp120 core in its different functional states, a full understanding of the function involved in gp120-CD4 interaction requires an analysis of molecular motions that drive complex formation. The conformational changes involve range from very subtle, local changes to global conformational changes that involve motions of significant amplitude for large parts of a protein. Dynamics plays an important role not only in the functional native state of many proteins, but also in the mechanism by which a protein reaches its functional conformation, e.g., the protein folding process, is a highly dynamic process. In the case of the HIV-1 gp120, the dynamics is of crucial importance in the process of conformational transition from the CD4-unliganded state to the CD4-bound state due to unusually large structural differences between these two states. Although there is currently no experimental technique that allows monitoring of protein conformational changes at atomic resolution, it has been found that further insights into the dynamic structure/function relationship of a protein can be gained through analysis of the computationally simulated dynamics [17]. For example, previous study on the dynamic properties of gp120 using a 10 ns molecular dynamics (MD) simulations [18] has demonstrated that CD4 binding reduced conformational flexibility in certain loops of the outer domain. The other MD studies indicated that CD4 binding partially locked the bridging sheet in gp120 but leaved the β2-β3 ribbon flexible [19, 20]. A shortcoming of the MD approach, however, is that currently available computing power permits only short simulations to be run for biological macromolecules, i.e., typically of the order of tens of nanoseconds. This time scale may insufficiently explore the configurational space available to a protein, and is a few orders of magnitude smaller than that on which most biological processes take place. An alternative approach to MD is to generate an ensemble of structures randomly without using Newton's equations of motion. CONCOORD [21] takes this approach as a fast way to generate structural ensemble that explores configurational space more fully. Previous studies [21–25] have shown that ensemble generated by CONCOORD is particularly useful in investigating global motions that a protein is able to perform around an equilibrium position.

In this study, we have performed comparative modeling technique to generate the three-dimensional models of gp120 in its different conformational states with modeled V3 and V4 loops, i.e., the gp120-CD4 complex, gp120 alone but in the CD4-bound state (CD4-free gp120), and gp120 in the CD4-unliganded state (unliganded gp120). These homology models we subjected to CONCOORD computer simulation to generate respective ensembles. One of our aims was to characterize the molecular motions of gp120 in the three conformational states at global level. The other aims were to identify subdomains (or rigid structural units) in gp120 and to study differences in structural flexibility between the three states of gp120.

Materials and methods

Model constructions

The sequence of the HIV-1 HXBc2 isolate gp160 precursor (Swiss-Prot accession number P04578) was obtained from Swiss-Prot protein sequence database [26]. The amino acid sequences for the transmembrane glycoprotein gp41, 52 and 19 residues from the N- and C-termini of gp120, and the V1/V2 loop (Gly-Ala-Gly substitutes for 67 V1/V2 loop residues) were removed. The final sequence consists of residues numbered from 83 to 492, including sequences for the V3 and V4 loops. Structural templates of the gp120 core and the V3 loop were obtained from PDB protein structure database [27]. PDB entries 1G9M (chain G) [11] and 2BF1 [14] were used as templates for modeling the CD4-bound and CD4-unliganded gp120 cores, respectively. PDB entry 1CE4 [28] was used for modeling the V3 loop. For HIV-1 gp120 structure 1G9M, the D1D2 domains of CD4 (chain C), neutralizing antibody Fab 17b (chains L and H), sugar groups, and crystal waters were removed and only the gp120 core (chain G) was preserved. For the SIVmac 32H gp120 structure 2BF1, sugar groups were removed and only the SIV gp120 core structure was retained. Since the SIV gp120 core shares share 35% sequence identity and over 70% sequence similarity with the HIV-1 HXBc2 gp120 core, the SIV core structure can serve as an effective template to generate medium-accuracy models of the unliganded HIV-1 gp120 core [29]. The NMR V3 structure (PDB entry 1CE4, residues 296–331) was manually orientated towards the V3 loop base residues Gly-Ala-Gly of the gp120 cores from 1G9M (CD4-bound) and 2BF1 (unliganded) by visual inspection using the VMD program [30], respectively. These two appropriately orientated V3 loop coordinates were saved as templates for V3 loop modeling.

To obtain homology model of gp120 with V3 and V4 loops in the CD4-bound conformational state, first, the sequence alignment between the templates of the V3 loop (orientated 1CE4) and the gp120 core (amended 1G9M) and the target sequence of the gp120 was constructed by an in-house written "alignment" script using the MODELLER software package [31]. The obtained sequence alignment was shown in Fig. 1. Second, 20 gp120 models were generated based on the template structures of gp120 core and orientated V3 loop using an in-house written script "get-model" for the MODELLER. The V4 loop coordinates were generated and optimized by a loop modeling subroutine within the "get-model" with a 3D_INTERPOLATION algorithm and a thorough optimization protocol, respectively. Finally, these 20 models were clustered and an optimized average model was generated by a MODELLER script "cluster". The transferred coordinates for a given target atom are the average of the 20 models in the largest cluster of corresponding atom with a cluster cut off of 1.5 Å. The final gp120 model was in the CD4-bound state without CD4 presented, which was referred to as the CD4-free (or CD4-bound) gp120 in the rest of the text.

To obtain the model of gp120 in complex with CD4, the CD4-free model was superimposed onto the crystal structure of the complex (1G9M) to obtain a relative orientation with respect to CD4. The D2 domain of CD4 was removed but the CD4-free gp120 and CD4 D1 domain taken from the crystal complex were preserved. The final model was referred to as the gp120-CD4 complex in the context. The gp120 within the gp120-CD4 complex was referred to as the CD4-complexed gp120 in the rest of the text.

To obtain homology model of gp120 with the V3 loop in the unliganded state, we have to initially obtain the unliganded HIV-1 gp120 core model from its template, the SIV unliganded gp120 core. The optimized unliganded HIV-1 gp120 core model and the orientated V3 loop were subsequently used as templates to model the unliganded HIV-1 gp120 with the V3 loop. The procedure of the sequence alignment, model generation and structure cluster and optimization resembles that of generating CD4-bound model. The final optimized average model in the CD4-unliganded state was referred to as the unliganded gp120 in the rest of the text.

Generation of structural ensembles

The three models, gp120-CD4 complex, CD4-free and unliganded gp120, were subjected to CONCOORD [21] simulation to generate respective ensembles. CONCOORD works in two phases. In the first phase, it derives a table of upper and lower distance constraints for various inter-atomic distances (covalent, ionic, and hydrogen bond etc.). These distance limits depend on distances measured in the parent structure and the strength of the inter-atomic

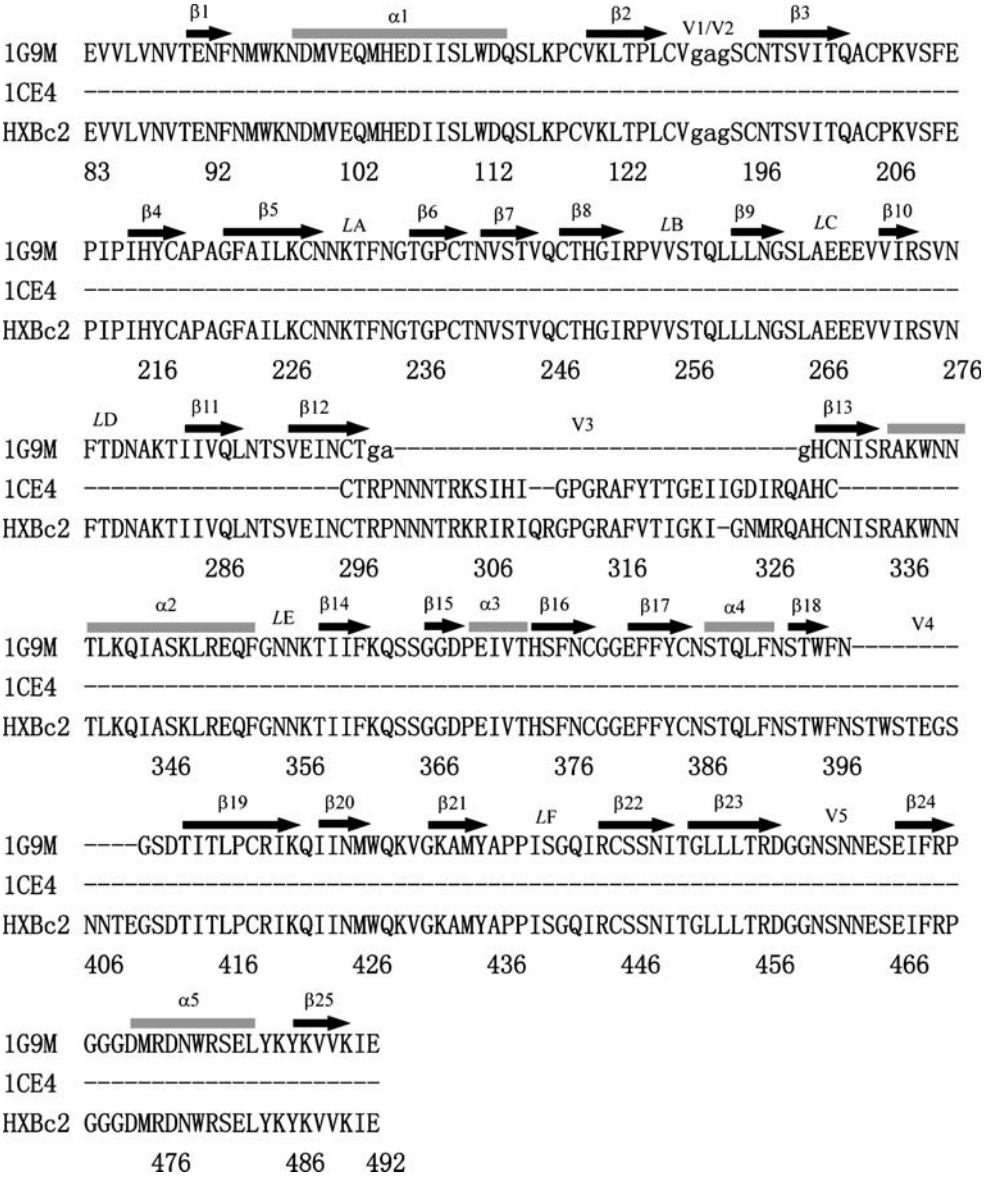

Fig. 1 Multiple sequence alignment between templates of the gp120 core (PDB entry 1G9M, Chain G) and V3 loop (PDB entry 1CE4), and the target sequence of the HXBc2 isolate (Swiss-Prot accession number P04578). The secondary structure elements are assigned according to the x-ray core structure with arrows for β-strands and rods for α-helices. The "gag" sequence in the V1/V2 and V3 loops of the gp120 core is the consequence of the truncation

interaction. In the second phase, CONCOORD generates an ensemble of structures that fulfills the inter-atomic distance bounds for all pairs of atoms. Starting from random coordinates, corrections are applied iteratively to the positions of those pairs of atoms that do not satisfy their specified distance constraints. This process is repeated until all constraints are satisfied. The ensemble generated in this way can be thought of as a set of uncorrelated structural snapshots, with the temporal order of frames scrambled. In many ways, however, the CONCOORD ensemble can be treated and analyzed in exactly the same way as a trajectory of snapshots produced by a MD simulation [21–25].

In this study, the following protocol was used to generate ensembles using CONCOORD. For generation of distance limits the DIST program was run as follows: the secondary structure-related atom pair recognition was performed using the DSSP program [32]; alternative non-bonded interactions were allowed; cut-off radius for non-bonded interacting pairs was left at 4.0 Å default value, as were all other analysis options. For structure generation the DISCO program was run as follows: the maximum number of iterations per structure and maximum number of trials per run were adjusted at 5000 and 500, respectively, to ensure convergence; chirality was checked "on the fly"; and all the other analysis options were left at default values. CONCOORD was run 3 times, once for each of the models generated by MODELLER. 500 conformations were finally generated for the gp120-CD4 complex, CD4-free gp120, and unliganded gp120, respectively.

Essential dynamics analysis

Essential dynamics (ED) method [33] is a powerful tool for filtering large-scale concerted motions from an ensemble of structures. The method yields the directions in configura-

tional space and is related to principal component analysis and quasiharmonic analysis [33–36]. ED is based on the diagonalization of the covariance matrix built from atomic fluctuations in a trajectory from which the overall translation and rotation have been removed. The covariance matrix C between two cartesian atomic coordinates x_i and x_j is defined as:

$$C_{ij} = < (x_i - < x_i >)(x_j - < x_j >) > \quad (1)$$

Where the <(quantity)> denotes an ensemble average. Applied to ensembles obtained from CONCOORD, we derive a matrix Λ by diagonalizing C_{ij} with a transformation matrix T:

$$\Lambda = T^T C T \quad (2)$$

The columns of T are the eigenvectors v_i, which describe collective modes of motions, with the first column being the most significant motion. The diagonal elements of Λ are the eigenvalues, which describe magnitudes of the motions along corresponding vectors. The required covariance matrix, eigenvectors, and eigenvalues were calculated using the g_covar program within the GROMACS package [37, 38]. Only backbone atoms were included because it has been shown that this simplification best detects the large-scale concerted motions in proteins [33, 39, 40].

The DYNAMITE program [23] was utilized to generate "porcupine plot" that was able to provide a graphical representation of the motion held in an eigenvector v_i. For instance to visualize eigenvector 1, a cone is drawn for each residue, starting from the C_α atom and projecting in the direction of component of the first eigenvector that corresponds to that residue. In addition, the length of the cone represents the motional amplitude of that atom. Given the eigenvectors of the ED decomposition, a script was generated by DYNAMITE to allow the molecular graphics program VMD [30] to automatically plot these cones onto the protein.

Covariance web analysis

A matrix in which each pair of atoms has an associated per-atom normalized covariance (PANC) can be calculated from the covariance matrix C using:

$$C'ij = \frac{Cxixj + Cyiyj + Czizj}{|(Cxixi + Cyiyi + Czizi)(Cxjxj + Cyjyj + Czjzj)|^{1/2}} \quad (3)$$

Where C'_{ij} is the PANC of atoms i and j and x_i, y_i, z_i, x_j, y_j, and z_j are the Cartesian coordinates of atoms i and j. C'_{ij} takes a value from -1 to +1 so that a value of 1 means that atoms i and j move together, a value of -1 means that they move in opposite directions and a value of 0 means that the atomic movements are uncorrelated. An interesting way to interpret this resulting matrix is to draw a line on a three-dimensional representation of the protein to connect any two atoms i and j with a correlation coefficient for their motions above a given threshold (typically 0.7) [22, 23], which yields an image in which correlated regions are linked by a web, as if their motions were constrained by a network of elastic rods. This approach can be used for identifying rigid domains or subdomains in proteins. The covariance web plots for the gp120-CD4 complex, CD4-free, and unliganded gp120 were generated by using DYNAMITE [23] and VMD [30].

Calculation of the B-factor

The crystallographically determined temperature factor (B-factor) provide one of the few measures of protein flexibility, and therefore simulation studies often attempt to reproduce them in order to test the verisimilitude of the dynamics observed [41, 42]. B-factor can be calculated from the mean square fluctuation (MSF) using the following expression:

$$B = \frac{8\pi^2}{3}(MSF). \quad (4)$$

The B-factors for the CD4-complex, CD4-free and unliganded gp120 were calculated from the MSF of CONCOORD simulations, respectively.

Results

Descriptions of models of gp120-CD4 complex, CD4-free and unliganded gp120

Both of the optimized HIV-1 gp120 models in the CD4-free and unliganded forms are composed of residues 83–127 and 195–492, summing up to a total of 3471 atoms. The V3 and V4 loops were modeled on the gp120 cores, while the V1/V2 loops are absent. A ribbon diagram of the CD4-free model was shown in Fig. 2a. This model closely resembles its template, the gp120 core from 1G9M. The superposition of the two structures gives a backbone root mean square deviation (RMSD) value of 0.36 Å. The overall structure can be divided into two major domains, the inner domain and the outer domain. The inner domain includes the N- and C- termini, a two-helix, a small two-stranded bundle, a six-stranded β-sandwich at its termini-proximal end and a projection at the distal end from which the V1/V2 stem (β2-β3 ribbon) emanates. The outer domain is a stacked double barrel that lies alongside the inner domain so that the outer barrel and inner bundle axes are approximately parallel. The V3 loop lies beneath the distal end of the outer

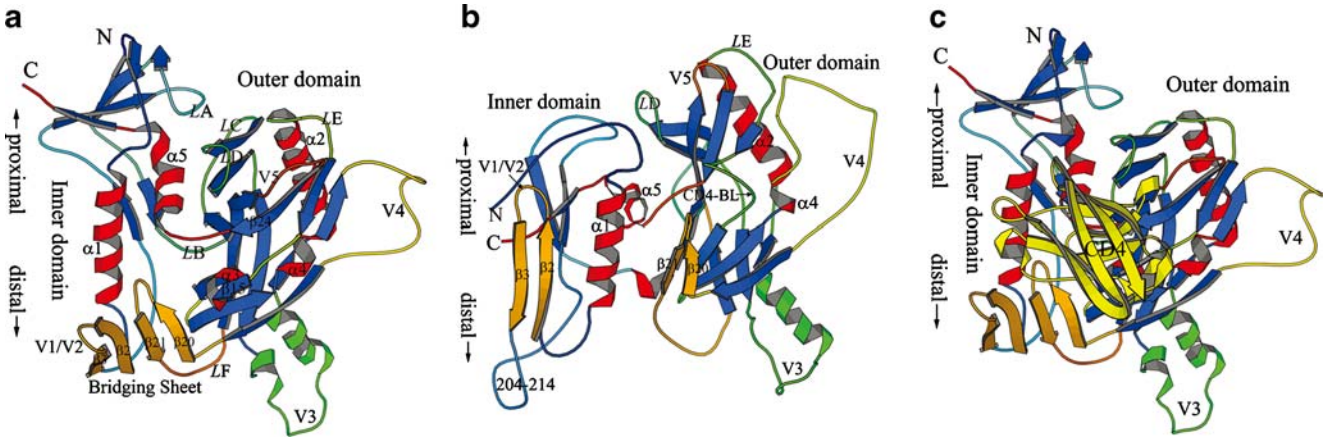

Fig. 2 Structures of the gp120 homology models. **a** Ribbon diagram of the CD4-free gp120. The α-helices, β-strands, bridging sheet and V3 loop are colored in red, blue, orange and green, respectively. **b** Ribbon diagram of the unliganded gp120. The structural elements are colored in the same color as in **a**. **c** Ribbon diagram of the gp120 in complex with CD4. The structural elements within the gp120 are colored in the same color as in **a** and **b**; and CD4 molecular are colored in yellow

domain, which acts as a connection between β12 and β13 and emanates from the body of the outer domain. The distance of the highly conserved Pro-Gly at V3 tip to the V3 base is ~ 30 Å. Although certain conformational differences are observed in V3 loops between our model and the most recently determined V3-containing gp120 core [43], the extended nature of the V3 in our model is in agreement with the crystal structure. The V4 loop extends away from the right side of the outer domain and adopts an open conformation. Another minidomain distinguishing from the inner and outer domains is the "bridging sheet", which stands below the distal ends of both the inner and outer domains. The V1/V2 stem emanating from the inner domain and the β20-β21 emanating from the outer domain constitute the antiparallel, four-stranded bridging sheet.

Figure 2b shows the unliganded HIV-1 gp120 model, which is similar to its template, the unliganded SIV gp120 core [14]. The RMSD between the common backbone atoms of them is 0.82 Å. The unliganded gp120 model also has two major domains as described for the CD4-free gp120. The inner domain consists of the N- and C-termini, a three-stranded β-sheet at its termini-proximal end, V1/V2 stem (β2-β3 ribbon), α-helix α1, and a short α-helix α5 at the inner-outer domain junction. The connecting segment (residues 218–228) between the β3 of the V1/V2 stem and the β5 of the three-stranded β-sheet was not resolved in the crystal structure of the SIV gp120. This segment corresponds to a segment comprising residues 204–214 in the unliganded HIV-1 gp120, which was modeled and optimized by using loop modeling sub-routine within MODELLER. As shown in Fig. 2b, the segment 204–214 presents a disordered loop conformation, extending away from the V1/V2 stem base of the unliganded HIV-1 inner domain (Fig. 2b). The outer domain has, with some local exceptions, globally the same structural organization as that in the CD4-free gp120. The V3 and V4 loops form protruding excursions on the body of the outer domain as they do in the CD4-free gp120. The major differences in the outer domain between the CD4-free and unliganded models are: i) the lengths in loops, i.e., the V4, V5, *L*E are slightly longer in the unliganded form; ii) the connection between β14 and β16 adopts a extended loop conformation in the unliganded gp120, which can be regarded as the "CD4-binding-loop" (CD4-BL) because it moves when CD4 associates and presents a extended strand (β15) and an α-helix (α3) in the CD4-bound conformation, with the β15 forming main-main hydrogen bonding with CD4's C" strand. The bridging sheet presented in the CD4-free gp120 is absent in the unliganded gp120. Although the two β ribbons, V1/V2 stem and β20-β21 that can form a bridging sheet is ordered, a space of 22−29 Å intervenes between them. This leads to the speculation that this minidomain can not form unless the receptor and/or coreceptor bind to gp120.

The model of the gp120-CD4 complex was shown in Fig. 2c. The structural organization of the CD4-complexed gp120 is identical to that of CD4-free model except that CD4 D1 domain has been docked onto it.

In the CD4-free gp120 model, the coreceptor CD4 binding sites are composed of the β20-β21 ribbon, V1/V2 stem, loops *L*D, *L*E, V5, β15-α3 excursion, and β24-α5 connection. All these structural components from the inner domain, the outer domain, and the bridging sheet constitute an unusually large CD4 binding cavity (Fig. 2a). The CD4 Phe43 binding pocket is located at the center of the large cavity. Specifically, it lies at the intersection of the inner domain, the outer domain, and the bridging sheet, and is deeply buried, extending into the hydrophobic interior of gp120. Residues lining the Phe43 pocket are primarily from the β20-β21 ribbon, β15-α3-β16, β24-α5 connection, and

β8-β9 connection (*L*B). However, neither the receptor CD4 nor the coreceptor CXCR4 or CCR5 site is properly formed in the unliganded model (Fig. 2b). As shown in Fig. 2b, the CD4-BL projects away from the center of the outer domain; and the β20-β21 ribbon, one half of the bridging sheet, stands beneath it. Together, the α1, α5, β20-β21 ribbon, and CD4-BL create a long, narrow channel at the intersection surfaces of the inner and outer domains. Apparently, the binding cavity for CD4 and binding pocket for CD4 Phe43 are absent in the unliganded state and can not be formed unless induced by CD4 binding.

ED analysis

The CONCOORD structural ensembles of the three gp120 models were subjected to ED analysis. In all cases only a few eigenvectors were found with significant eigenvalues. The eigenvalues as a function of eigenvector index were shown in Fig. 3. The trace of the covariance matrix after diagonalization, which describes the total MSF of the ensemble, is 59.9, 78.2, and 79.6 nm^2 for the CD4-complexed, CD4-free and unliganded gp120, respectively. For the three ensembles, the first 4 eigenvectors contribute 66.0%, 70.2%, and 70.8% to the total MSF, and the first 10 eigenvectors contribute 84.5%, 87.8%, and 84.7% to the total MSF, respectively. These results indicate that the first 10 eigenvectors, especially the first 4 eigenvectors, emerge with appreciable freedom.

Collective molecular motions

Figure 4 shows, in porcupine representation, the large-scale collective motions of the gp120-CD4 complex and the

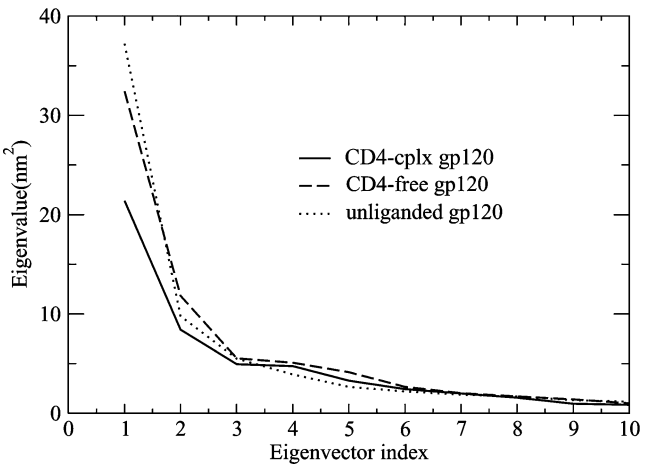

Fig. 3 Eigenvalues as a function of eigenvector index obtained from essential dynamics analyses of CONCOORD ensembles of the CD4-complexed, CD4-free, and unliganded gp120. Only the first 10 largest eigenvalues are shown out of 3114. The CD4-complexed gp120 is abbreviated to CD4-cplx gp120

CD4-free and unliganded gp120. In the case of the CD4-complexed gp120 (Fig. 4a), eigenvector 1 describes a rotation/twisting motion between the inner domain and the outer domains. Specially, the N-, C-termini and the termini-proximal β-bundle in the inner domain, and the V3 and V4 loops in the outer domain show the largest fluctuations as they show the longest "quill"; the V1/V2 stem in the bridging sheet and the loops *L*C, V5, and *L*E in the outer domain show moderate fluctuations; the α1 and α5 in the inner domain and certain structural components (such as β20-β21 ribbon, α3-β15 etc) in the outer domain contacting with or close to CD4 show the smallest fluctuations as they have the shortest "quill". The motion of CD4 resembles that of the gp120 inner domain. Briefly, CD4 and bridging sheet, together with the inner domain, perform a twist around an axis running from the outer domain to the inner domain with respect to the outer domain, despite almost uniformly moderate fluctuations occurring across the entire CD4 molecule. The second largest collective motion (described by eigenvector 2) of gp120-CD4 complex is illustrated in Fig. 4b. This motion can be described as a rotation/twisting motion of the inner domain (in particular the N-, C-termini), bridging sheet, and V3 loop relative to CD4 and large parts of the outer domain. The third dominating motion (described by eigenvector 3) consists of a twist of the N-, C-termini and V3 loop relative to CD4 and outer domain. As shown in Fig. 4c, the outer domain and bridging sheet, together with CD4, are dominated by a vortex that rotates around an axis running from the center of the outer domain to CD4. Another vortex formed by the N, C-termini and V3 loop rotates around the axis in the opposite direction. The fourth ranked motion is dominated by an elongation of the inner domain and outer domain along respective major axes, resulting in a thinning of the two domains (data not shown).

In terms of the CD4-free gp120, the largest collective motion described by eigenvector 1 was found to be a twist of the inner domain relative to the outer domain (Fig. 4d), i.e., the two domains rotate in opposite directions around an axis connecting the centers of the two domains. This mode resembles the twisting motion described by eigenvector 1 for the CD4-complexed gp120, whereas the rotation directions of the inner domain and outer domain are reversed compared to those of the inner and outer domains of the CD4-complexed gp120. The second largest collective motion of the CD4-free gp120 corresponds to a twisting motion between the two domains. These to domains rotate around an axis running roughly through the centers of them, with a large part of the outer domain (especially the proximal end and V4 loop) forming a vortex, rotating in opposite direction relative to a vortex formed by the N-, C-termini, bridging sheet, and V3 loop (Fig. 4e). This motion is similar to the second ranked motion of the CD4-

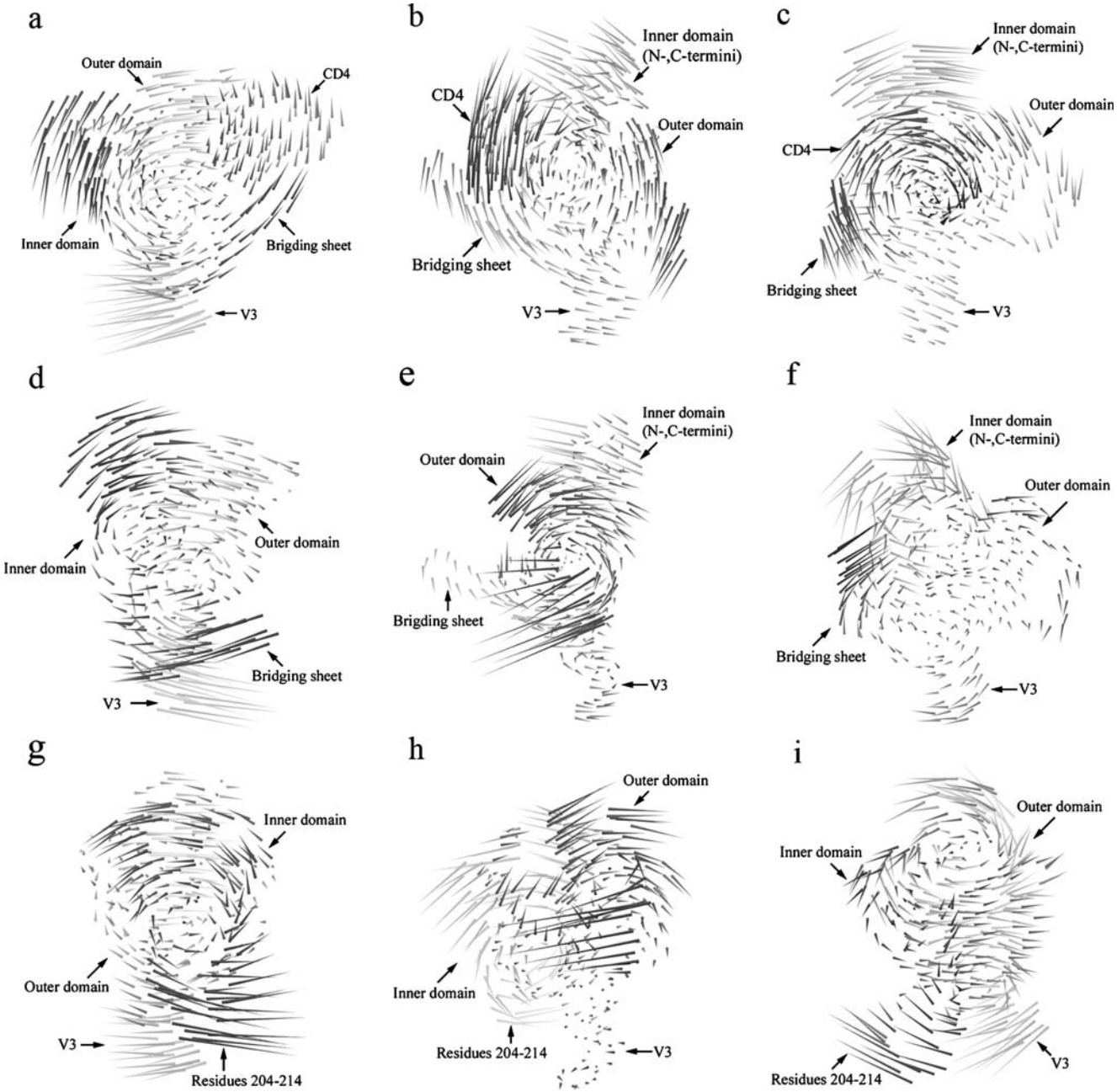

Fig. 4 Porcupine plots of the principal modes of molecular motions calculated from CONCOORD ensembles of the gp120-CD4 complex, CD4-free, and unliganded gp120. **a** The most significant motion of the gp120-CD4 complex is dominated by a twist of CD4 and the inner domain and bridging sheet of gp120 relative to the outer domain of gp120. The view is from the inner domain to the outer domain. **b** The second ranked motion of the gp120-CD4 complex is dominated by a twist of CD4 and a large part of the outer domain relative to the inner domain, bridging sheet and V3 loop. The view is from the outer to inner domains. **c** The third ranked mode of the gp120-CD4 complex describes a twist of the N-, C-termini and V3 loop relative to CD4, bridging sheet and a large part of the outer domain (without including the V3 loop). This view is looking "down" from CD4 molecule. **d** The most significant motion of the CD4-free gp120 is a twist of the inner domain (including the bridging sheet) relative to the outer domain. The view is from the inner to outer domains. **e** The second ranked motion of the CD4-free gp120 is dominated by a twist of a large part of the outer domain (without including V3 loop) relative to the inner domain, bridging sheet, and V3 loop. The view is from the outer to inner domains. **f** The third ranked mode of the CD4-free gp120 describes a flexing of the N-, C-termini within the inner domain relative to a clockwise vortex formed by the bridging sheet and outer domain. This view is from the CD4 binding cavity. **g** The most significant motion of the unliganded gp120 is dominated by a twist of the inner domain with respect to the outer domain. This view is from the inner to outer domains. **h** The second ranked motion of the unliganded gp120 is dominated by a twist of a large part of the outer domain relative to the inner domain, with the β20-β21 and V3 in the outer domain exhibiting the smallest fluctuations. The view is from the outer to inner domains. **i** The third ranked mode of the unliganded gp120 describes a flexing of residues 204–214 and V3 loop relative to the anticlockwise vortex formed by large parts of the inner and outer domains. This view is from the inner to outer domain

complexed gp120. Eigenvector 3 describes a relatively large vortex formed by the bridging sheet and the entire outer domain that rotates around an axis running through the Phe43 pocket, resulting in a flexing of the vortex relative to the N-, C-termini of the inner domain (Fig. 4f). The fourth eigenvector characterizes a shortness of the inner and outer domains along respective major axes, resulting in a condensation of the two domains (data not shown).

In the case of the unliganded gp120, the motion described by eigenvector 1 is a twist between the inner and outer domains. As shown by the porcupine plot (Fig. 4g), the two domains form a pair of opposite spin vortices that rotate around an axis connecting the centers of the two domains. The two domains rotate in the same direction as the twisting motion described by eigenvector 1 for the CD4-complexed gp120 (Fig. 4a), but in the opposite direction compared to the twisting motion described by eigenvector 1 for the CD4-free gp120 (Fig. 4d). Eigenvector 2 of the unliganded gp120 describes a twisting motion between the inner and outer domains. These two domains form two vortices that rotate in opposite directions around an axis running through the centers of the two domains (Fig. 4h). This mode is similar to the second largest modes obtained from the CD4-complexed and CD4-free ensembles, although the smallest fluctuations were observed in the β20-β21 ribbon, *L*F, and V3 loop of the unliganded gp120. We also note the long, narrow channel at the interfaces of the inner and outer domains undergoes a slight closing motion. Unlike the third ranked motions for the CD4-complexed and CD4-free gp120, the eigenvector 3 of the unliganded gp120 is dominated by a single anticlockwise rotation vortex formed by a large part of the inner and outer domains, rotating around an axis running from the outer domain to the inner domain. Other parts of the unliganded gp120 such as residues 204–214 in the inner domain and the V3 loop in the outer domain undergo flexing motion in the opposite direction relative to the single vortex (Fig. 4i). Eigenvector 4 describes an elongation of the inner and outer domains, resulting in respective thinning of the two domains (data not shown).

Covariance web analyses and subdomain

As described in the Methods section, rigid domains or subdomains of the gp120-CD4 complex, CD4-free and unliganded gp120 can be identified by their tendencies to undergo collective motions. We have arbitrarily chosen a threshold of 0.7 to draw web lines on each of the three graphical representations of gp120.

Figure 5 shows the covariance web plots for the three gp120 models. As would be expected, a mesh of lines interconnects the atoms not only within secondary structure elements, consistent with locally correlated motions of atoms within these elements, but also between different structural components, predicting higher order correlations between them. The structural components moving in concert with each other can be identified as the rigid structural unit or subdomain. In the case of the gp120-CD4 complex, the covariance web analysis of the CONCOORD ensemble has identified a total of 5 domains that tend to move with high internal correlation (Fig. 5a): 1) the gp120 N-, C-terminal domain (including the β-bundle at the termini-proximal end of the inner domain); 2) the gp120 bridging sheet domain; 3) the V3 domain; 4) a region composed of the *L*C, *L*D, *L*E, V5, α2 and β-bundle that are located at the proximal end of the outer domain; and 5) CD4 D1 domain. Figure 5a also shows that concentrated web

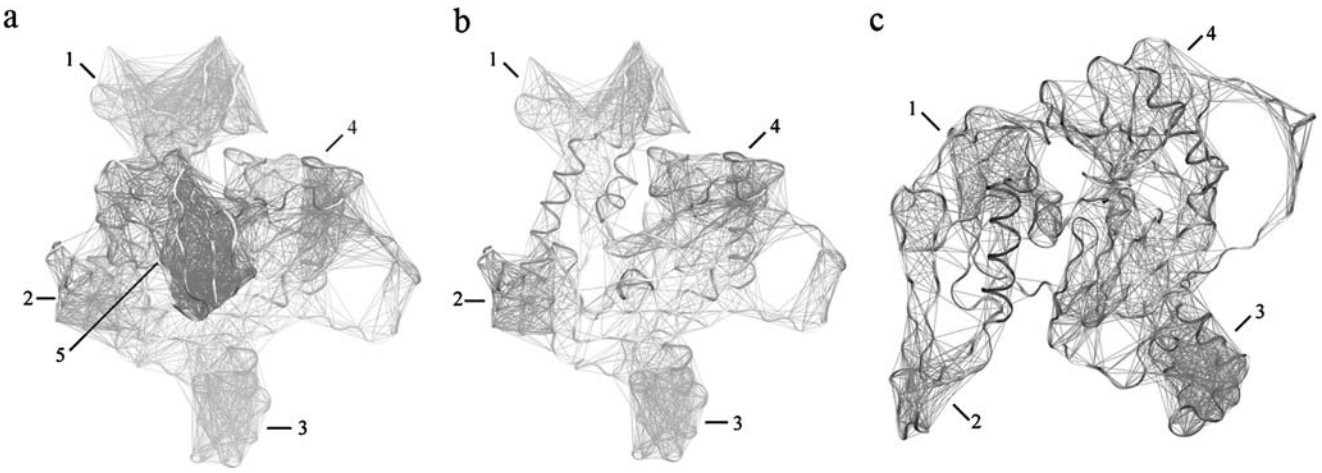

Fig. 5 Covariance web plots for **a** the gp120-CD4 complex, **b** the CD4-free gp120, and **c** the unliganded gp120. These plots reveal correlated motions calculated from the CONCOORD ensembles. The lines represent C_α atoms related by a PANC value greater than 0.7 and highlight 5, 4, and 4 rigid regions (enumerated in the main text) that move together in the gp120-CD4 complex, CD4-free and unliganded gp120, respectively

lines covered across the overall CD4 structure, indicating that a rigid core is formed in CD4. In addition, the web lines that link CD4 with several gp120 structural components such as the bridging sheet, *L*B, β15-α3, *L*D, and V5 were also observed. This means that the movement of CD4 appears to correlate closely with the movements of these gp120 components. Figure 5b shows the covariance web plot for the CD4-free gp120, in which the rigid subdomains identified are similar to those of the CD4-complexed gp120. For instance there were four apparent subdomains: 1) the N-, C-terminal domain; 2) the bridging sheet domain, 3) the V3 loop domain; and 4) the region that is located at the proximal end of the outer domain. Other components such as the α1, α5, β15-α3, and V4 appear to be independently mobile units because few web lines connect between them. The covariance web plot of the unliganded gp120 (Fig. 5c) also reveals 4 major domains. They are 1) the N-, C-terminal domain (including the N-, C-termini and the β-bundle at the proximal end of the inner domain); 2) the linker region (residues 204–214) between the V1/V2 stem and the β-bundle at proximal end of the inner domain; 3) a region that is composed of the V3 loop and a part of the outer domain at its distal end; 4) a region that is located at the proximal end of the outer domain. The other structural components such as the V1/V2 stem, α1, β20-β21 ribbon, CD4-BL, and V4 are independent rigid units because the concerted atom fluctuation occurs almost within respective components. The differences in rigid domain between the CD4-free, CD4-complexed, and unliganded gp120 are: i) the bridging sheet in the CD-free and CD4-complexed gp120 moves with high internal correlation, while the two halves of the bridging sheet, V1/V2 stem and β20-β21 ribbon in the unliganded gp120, act as the independent rigid unit; 2) the V3 loop in the CD4-complexed and CD4-free gp120 appears to be a separate domain because the motion of V3 seldom correlates with its neighboring components. On the contrary, the motion of the V3 loop in the unliganded gp120 appears to associate closely with the motion of the structural components that are located at the distal end of the outer domain, leading to a formation of a larger rigid domain compared to the V3 domain in the CD4-free and CD4-complexed gp120.

Comparison of the B-factor

Figure 6 shows the crystallographically determined B-factors for the CD4-complexed gp120 (PDB entry 1G9M, Chain G) and the B-factors calculated from CONCOORD simulations of the CD4-complexed, CD4-free and unliganded gp120. The experimental B-factors are generally smaller than the calculated B-factors because the simulations sampled far more conformations of gp120 than crystallography. Although a quantitative comparison between the simulated and experimental B-factors may not be applicable due to the differences between the experimental and simulation conditions [44], the calculated B-factor does not detract from the qualitative agreement with the experimental values. For example for much of the structure, those regions that present increased B-factors in the crystal structure also show increased B-factors in the ensemble of simulated structures and this relationship also holds true for the regions with decreased B-factors. This indicates the computer simulation predicts the identity of mobile regions better than it predicts the degree of their mobility. Here we primarily focused on the comparison of the B-factors calculated from CONCOORD simulations to identify high-mobility regions and the differences in B-factor between models of the CD4-complexed, CD4-free, and unliganded gp120.

For the CD4-complexed and CD4-free gp120, Fig. 6 shows that their simulated B-factor curves match well with

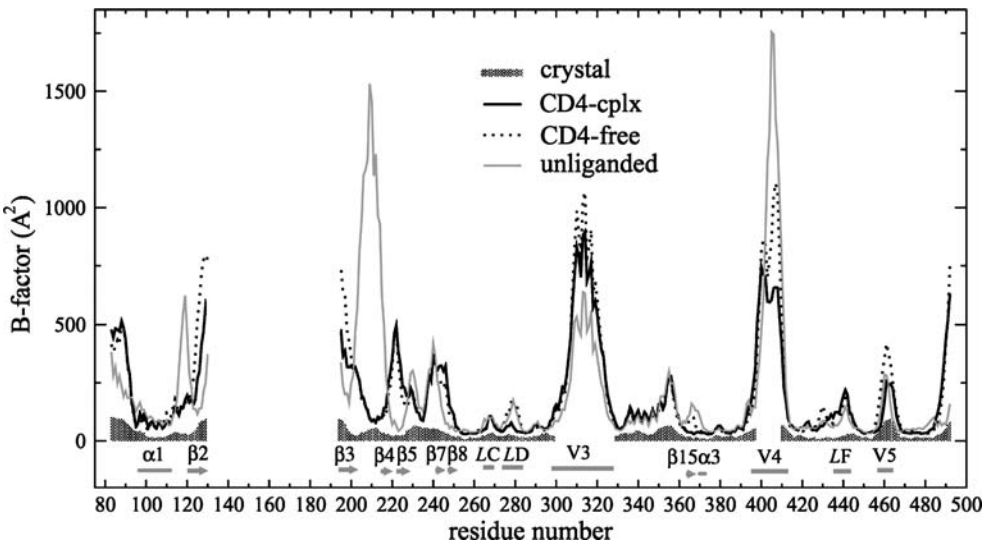

Fig. 6 Comparison of gp120 Cα B-factors calculated from RMS fluctuations of CONCOORD ensembles between the CD4-complex (solid line), CD4-free (dashed line), and unliganded (gray line) states. The experimentally determined B-factors of the CD4-bound gp120 (PDB entry 1G9M) are shown in gray area. Note that the experimental B-factors of the loops V1/V2, V3, and V4 are missing. The CD4-complexed gp120 is abbreviated to CD4-cplx gp120 and the secondary structure elements and loops V3 and V4 were marked according to the crystal structure of the CD4-bound gp120

each other with the exception of certain regions such as the V1/V2 stem (β2-β3), β15-α3, β20-β21 ribbon's connection, and some variable loops such as the *L*D, V3, V4, V5. These regions show higher B-factors in the CD4-free gp120 than in the CD4-complexed gp120, indicating that the removal of CD4 increases the fluctuations of these regions. Specially, the flexibility of the β2-β3, *L*D, and β15-α3 increases essentially upon removal of CD4. This is not surprising, because CD4 does not tether these components anymore.

Comparison of the B-factors of the CD4-complexed gp120 with those of the unliganded gp120 also shows that there are many features common to both curves (Fig. 6). However, several apparent differences were noted. For example, the B-factors of the N-, C-termini, β2-β3 ribbon, β4-β5 and their connection, V3, and *L*F are significantly lower in the unliganded gp120 than in the CD4-complexed gp120, indicating that these regions are relatively stable in the unliganded state. On the contrary, the connection between the α1 and β2, the connection (residues 204–214) between the β3 and β4, *L*D, CD4-BL (β15-α3), and V4 show higher B-factors in the unliganded gp120 than in the CD4-complexed gp120. CD4 binding is responsible for the decreased flexibility of the regions interacting directly with CD4 (such as *L*D, CD4-BL, and a part of the V5). The decreased B-factors for the α1-β2 linker and the β3-β4 linker in the CD4-complexed gp120 are probably caused by the reorientation of the V1/V2 stem (β2-β3 ribbon), which may drive the formation of the bridging sheet. Interestingly, in the CD4-complexed gp120, the increased flexibility of the V1/V2 stem, N- and C-terminal domain (which are composed of the N-, C-termini, β4-β5 and their connection, and β7-β8), and the loops V3 and *L*F could be related to a serial of functional tasks performed by gp120 upon CD4 association but prior to the membrane fusion.

Discussion

The prerequisite of our analysis of the gp120 dynamics is to obtain in depth knowledge of the structures in both the CD4-bound and unliganded states. The 100% sequence identity between the target sequence of the HXBc2 and its template 1G9M results in an accurate core structural model of the liganded gp120 (CD4-complexed or CD4-free gp120). The approximately accurate V3 and V4 loops were also modeled on this core. Structure validation using PROCHECK [45] indicates more than 98% of residues in this model fall within the favored regions of the Ramachandran plot. The direct experimental structure of the unliganded HIV-1 gp120 is not available at this stage. However, given the high sequence identity between gp120 derived from the SIV and HIV, the unliganded HIV gp120 is likely to assume a similar conformation with that of the SIV. It was noted [29] that it is possible to generate medium-accuracy unliganded HIV-1 gp120 models with ~90% of the main-chain atoms being modeled with RMS error less than 1.5 Å. In our unliganded gp120 model with the modeled V3 loop, there are more than 90% of residues that have appropriate dihedral angel distributions. Therefore we believe that the differences in conformation between homology models of the unliganded gp120 and the CD4-complexed/CD4-free gp120 should reflect the conformational changes induced by receptor binding. Also, these three models in their different states (CD4-complexed, CD4-free, and unliganded) are suitable for exploring the molecular motions and conformational flexibility of gp120.

To study the mobility and flexibility of a protein, simulation techniques such as the MD or CONCOORD are required to generate ensemble of a structure in which each frame contains an accessible configuration. Although the MD structures are in principle free to access all physically obtainable conformation, the relatively short timescales that this approach can currently access; the incremental search protocol that is used to obtain individual members of the ensemble; and the Brownian and viscous effects of the solvent environment, may limit the full exploration of the conformational space available to the protein [22]. On the contrary, the CONCOORD structures are derived from constraints that circumscribe a region of the conformational space that can be accessed. This space is thoroughly explored in a CONCOORD ensemble as a result of the highly randomizing protocol by which each member of the ensemble is generated [21]. Thus, the ensembles derived from the CONCOORD are most useful in identifying global large-scale motions that a protein is able to perform around an equilibrium position and in predicting the character of conformational changes.

For all three simulation systems, the ED analyses of CONCOORD ensembles reveal that the first 4 eigenvectors account for a large proportion of the total fluctuations. The modes of the large-scale collective fluctuations are probably involved in the mechanism of gp120-ligand (CD4, antibodies, or small molecular inhibitors) association and neutralization avoidance. Although the ED technique yields individual collective motions, all modes interplay with each other in a complex manner, and therefore, not necessarily all individual modes will correspond to a specific functional task. For purposes of simplicity and clarity, however, the individual modes described by the first 4 eigenvectors will be discussed in relation to the putative biological functions.

The most significant motions of gp120 in the CD4-complexed, CD4-free, and unliganded forms described by eigenvector 1 are all the twist of the inner domain relative to the outer domain. However, the variation of twisting directions among the three gp120 states may bring about

different functional effects. For the gp120-CD4 complex (Fig. 4a), the gp120 inner domain and bridging sheet, together with CD4 D1 domain, form an anticlockwise vortex, which rotates with respect to a clockwise vortex formed by the gp120 outer domain. The effect of such twist could grip CD4 molecule tightly because the bridge sheet and the components (such as the *L*D, V5, and *L*E) that are located at the proximal end of the outer domain move towards CD4. For the CD-free gp120 (Fig. 4d), the clockwise vortex formed by the inner domain and bridging sheet rotates relative to the anticlockwise vortex formed by the outer domain, resulting in a twisting mode in the opposite direction compared to that of the CD4-complexed gp120. This mode will enlarge the CD4 cavity because the bridge sheet and the components that are located at the proximal end of the outer domain move away from the cavity. For the unliganded gp120, the twisting motion is similar to that of the CD4-complexed gp120. The anti-clockwise vortex formed by the inner domain rotates with respect to the clockwise vortex formed by the inner domain (Fig. 4g). However, this twisting mode in the unliganded gp120 does not directly open or close the long, narrow channel.

All the second eigenvectors derived from the three simulation systems are dominated by a similar rotation/twisting motion of the outer domain with respect to the inner domain (Fig. 4b,e,h), although the bridging sheet and V3 loop in the CD4-complexed and CD4-free gp120 rotate in concert with the inner domain. For the CD4-complexed and CD4-free gp120s, this motion will enlarge the CD4-binding cavity and could be related to CD4 release, whereas for the unliganded gp120, this motion will slightly close the long, narrow channel. Of particular interest is that the V3 loop and *L*F in the CD4-complexed gp120 emerge with the largest motion among the three gp120 forms. This may be caused by CD4 binding and therefore we speculate that the high mobility of the *L*F and V3 would facilitate association of gp120 with the coreceptor CCR5 or CXCR4.

The rotations of the two vortices described by eigenvectors 3 for the CD4-complexed g120 will close the narrow part of CD4 cavity that is located at the interfaces (composed of the α5, *L*C, *L*D, β9, β10, and β11, see Fig. 2a) of the proximal ends of the inner and outer domains while open the Phe43 pocket that is located at the intersection of the outer domain, the inner domain, and the bridging sheet. This mode (Fig. 4c) could be related to the orientation and insertion of CD4 Phe43 into its binding pocket. However, the two vortices of the CD4-free gp120 rotate in the opposite directions (Fig. 4f) compared to those of the CD4-complexed gp120. This leads to the reversed effects on the narrow part of CD4 cavity and Phe43 pocket, i.e., enlarges the former and reduces the latter. The third eigenvector of the unliganded gp120 describes a flexing motion of the outer domain and inner domain relative to the V3 loop and the segment comprising residues 204–214 (Fig. 4i). This kind of flexing could influence reorientations of the V1/V2 stem, the V3 loop, and the connecting segment (residues 204–214) between the V1/V2 stem and three-stranded bundle at the termini-proximal end of the inner domain. In addition, such reorientations may also promote approach between the V1/V2 stem and the β20-β21 ribbon, leading to the formation of the bridging sheet.

The fourth eigenvectors of the CD4-complexed, CD4-free, and unliganded gp120 describe the elongation or shortness of the inner and the outer domains along respective major axes. These modes lead to the thinning or condensation of the two domains and may contribute to binding or release of CD4.

It is worth pointing out that the large-scale collective motions of gp120 in its three states do not represent explicit rotation/twisting or breathing (opening/closing) modes between the two domains as those have been observed in the T4 lysozyme [46], hyaluronate lyase [24], and cyclin-dependent kinase 2 [22]. Although the structures of gp120 in the CD4-bound and unliganded states can be divided into two major domains based on their structural organizations [11, 14], ED and DYNAMITE analyses demonstrate that certain structural components such as the bridging sheet, V1/V2 stem, and V3 loop can move concertedly with either the inner domain or the outer domain, resulting in a complicated combination of variable motional modes (rotation/twisting, opening/closing, or closing/flexing) of one mobile unit (or domain) with respect to the other mobile unit (or domain). We suggest that these perplexing motions reflect the peculiar conformational flexibility of the gp120, which could be related to the dynamics mechanism responsible for receptors/coreceptor association, and may also explain why it is hard for gp120 to induce potent neutralizing antibodies against CD4 and coreceptor binding sites.

Covariance web analyses reveal 4 major subdomains that move with high internal correlation in the three gp120 models. The other components such as the α1, α5, β15-α3 in the CD4-free gp120, and the V1/V2 stem, α1, β20-β21 ribbon, and CD4-BL in the unliganded gp120, appear to be independent rigid units. We infer from these results, that the association of the unliganded gp120 with CD4 is not a simple process of the one-step binding, but is a multiple-step induced fit involved not only the major subdomains, but also the other independent rigid components. It is not clear whether the bridging sheet has been formed upon CD4's association but prior to coreceptor or antibody 17b binding. MD studies [19, 20] indicated that CD4 binding just partially locked the bridging sheet but left the V1/V2 stem flexible. The covariance web plots derived from the

CONCOORD simulations also indicate that the α1 and V1/V2 stem are independent rigid mobile units. Thus, we conclude that the V1/V2 stem is not fully docked against the β20-β21 ribbon upon CD4 binding. The subsequent association of the coreceptor or antibody 17b, may facilitate conformational adjust of the V1/V2 stem and stabilize the CD4-bound state and lock the bridging sheet. We also note that the V3 loop in the unliganded gp120 moves in close concert with the structural components neighboring the V3 loop base (Fig. 5c), whereas the V3 loop in the CD4-complexed gp120 is an isolated rigid unit as it moves almost uncorrelated with other structural components (Fig. 5a). In addition, the V3 loop presents larger B-factors in the CD4-complexed gp120 than in the unliganded gp120 (Fig. 6). We speculate that the association of CD4 is responsible for both the "detachment" of the V3 loop from its adjacent components and the increased freedom of the V3 loop. The increased flexibility of the V3 in the CD4-bound state may facilitate the subsequent recognition and binding of the coreceptor CCR5 or CXCR4. Another interesting finding is that CD4 moves in concert with the bridging sheet, β15-α3, *L*D, and V5 in the gp120-CD4 complex. Such concerted motions underlie the profound effect of CD4's association on the orientation of these structural components, which might be the appropriate targets in the design of inhibitors interrupting CD4-gp120 interaction.

The B-factors of the three gp120 models calculated from the CONCOORD ensembles are larger than the crystallographically determined B-factors for the CD4-complexed gp120 (Fig. 6). This is probably caused by the simplified gp120 homology models and the search protocol that the CONCOORD used. For example, no inclusion of the sugar groups in models; a set of the simple geometric constraints; no Brownian and viscous effects of the solvent environment; and the highly-randomizing structure generation method, will lead to a full exploration of the conformational space that the protein can access. However, our simulations provide the qualitative consistency with the experimentally determined distribution of the mobile regions of gp120. Notably, the V3 and V4 loops in all the three simulated models and the segment comprising residues 204–214 in the unliganded gp120 are found to diffuse away from interaction with the rest of the molecule and adopt a highly flexible conformation. These results agree with the crystallographic data, indicating that these regions lack interpretable electron density in the crystals of the CD4-complexed and unliganded gp120. Interestingly, the association of CD4 with gp120 increases the B-factors for some regions, while decreases the B-factors for the other regions. In particular, CD4's association increases the flexibility of the V1/V2 stem (β2-β3 ribbon), *L*F, V3, and the N-, C-terminal subdomain that is composed of the N-, C-termini and β-bundle at the termini-proximal end of the inner domain. The increased flexibility of the V1/V2 stem, *L*F, and V3 may be crucial to the subsequent coreceptor association. It has been known that the gp120 is constrained in its assembly with gp41 into a trimeric complex on the viron surface [9, 10], where the N-, C-terminal subdomain of gp120 interacts with the gp41. The increased mobility of the N-, C-terminal domain caused by CD4 binding would then weaken the non-covalent gp120-gp41 interaction, and promotes the dissociation of gp120 from gp41. CD4's association decreases the mobility of the *L*D, CD4-BL, and a part of the V5 because of their contacts with CD4 in the gp120-CD4 complex. As shown in Fig. 6, the largest reduction in B-factors is found in the region of residues 204–214 and the connection between the α1 and β2. We speculate that CD4 binding information is transmitted by these regions to the N-, C-terminal domain, which could, as suggest above, promote the dissociation of gp120 from gp41, liberating the latter to commit to the fusion transition. For the β2-β3 ribbon, V3, V4, *L*A, *L*D, and CD4-BL (β15-α3), the high flexibility of these structural components might, on one hand, account for the failure of gp120 to induce potent neutralizing antibodies, on the other hand, modulate the kinetics and thermodynamics of interactions of gp120 with structurally dissimilar partners. Also worth noting is that the β2-β3 ribbon, which is one half of the bridging sheet, shows high flexibility (large B-factor) in all three gp120 states. Such a high flexibility may make it possible for the β2-β3 ribbon to adopt multiple conformational states in gp120. We therefore considered that some of the populated conformers near the unliganded or CD-bound conformation would serve as intermediates that might be the starting conformations favoring more efficient CD4 recognition/association under the physiological condition.

In summary, our homology models of the HIV-1 gp120 in the CD4-complexed, CD4-free, and unliganded forms represent three conformational states, namely the excited or pre-fusogenic state, the excited state in the absence of CD4, and the relaxed ground state, respectively. The CONCOORD computer simulations and subsequent ED analyses of the ensembles revealed that the large-scale concerted motions are dominated by intricately combinatorial rotations of the vortices formed between or within the inner domain, outer domain, bridging-sheet, and V3 loop. These modes will open/close the CD4 cavity and reposition certain structural components involved in ligand associations. We suggest that these perplexing motions provide one probable explanation for the failure of gp120 to induce potent neutralizing antibodies against CD4 and coreceptor binding sites. Covariance web analyses not only reveal the rigid subdomains but also identify certain components that move in concert with CD4, which may be suitable targets for structure-based inhibitor design as a potential route to

anti-HIV therapy. The comparative analyses of the simulated B-factors reveal changes in flexibility for certain structural regions, which are related either to CD4 association or to the subsequent dissociation of gp120 from gp41. We have also demonstrated the sophisticated plasticity of gp120, from its rigid core to a high-flexibility exterior coating that provides recognition specificity without compromising the capability to avoid neutralization.

Acknowledgements The authors thank High Performance Computer Center of Yunnan University for computational support. This work was supported by funds from a 973 project (2003CB415102) and partially supported by grants from Yunnan Province (2006C0008M, 07Z10756, 2007C163M) and Innovation Group Project from Yunnan University.

References

1. Barre-Sinoussi F, Chermann JC, Rey F, Nugeyre MT, Chamaret S, Gruest J, Dauguet C, Axler-Blin C, Vezinet-Brun F, Rouzioux C, Rozenbaum W, Montagnier L (1983) Science 220:868–871
2. Gallo RC, Salahuddin SZ, Popovic M, Shearer GM, Kaplan M, Haynes BF, Palker TJ, Redfield R, Oleske J, Safai B (1984) Science 224:500–503
3. Heeney JL, Hahn BH (2000) AIDS Suppl 14:125–127
4. Klein E, Ho R (2000) Clin Ther 22:295–314
5. Dalgleish AG, Beverley PC, Clapham PR, Crawford DH, Greaves MF, Weiss RA (1984) Nature 312:763–767
6. Feng Y, Broder CC, Kennedy PE, Berger EA (1996) Science 272:872–877
7. Trkola A, Dragic T, Arthos J, Binley JM, Olson WC, Allaway GP, Cheng-Mayer C, Robinson J, Maddon PJ, Moore JP (1996) Nature 384:184–187
8. Wu L, Gerard NP, Wyatt R, Choe H, Parolin C, Ruffing N, Borsetti A, Cardoso AA, Desjardin E, Newman W, Gerard C, Sodroski J (1996) Nature 384:179–183
9. Veronese FD, DeVico AL, Copeland TD, Oroszlan S, Gallo RC, Sarngadharan MG (1985) Science 229:1402–1405
10. Trkola A, Purtscher M, Muster T, Ballaun C, Buchacher A, Sullivan N, Srinivasan K, Sodroski J, Moore JP, Katinger H (1996) J Virol 70:1100–1108
11. Kwong PD, Wyatt R, Robinson J, Sweet RW, Sodroski J, Hendrickson WA (1998) Nature 393:648–659
12. Kwong PD, Wyatt R, Majeed S, Robinson J, Sweet RW, Sodroski J, Hendrickson WA (2000) Structure Fold Des 8:1329–1339
13. Wyatt R, Kwong PD, Desjardins E, Sweet RW, Robinson J, Hendrickson WA (1998) Nature 393:705–711
14. Chen B, Vogan EM, Gong H, Skehel JJ, Wiley DC, Harrison SC (2005) Nature 433:834–841
15. Kwong PD, Doyle ML, Casper DJ, Cicala C, Leavitt SA, Majeed S, Steenbeke TD, Venturi M, Chaiken I, Fung M, Katinger H, Parren PW, Robinson J, Van Ryk D, Wang L, Burton DR, Freire E, Wyatt R, Sodroski J, Hendrickson WA, Arthos J (2002) Nature 420:678–682
16. Myszka DG, Sweet RW, Hensley P, Brigham-Burke M, Kwong PD, Hendrickson WA, Wyatt R, Sodroski J, Doyle ML (2000) Proc Natl Acad Sci USA 97:9026–9031
17. Berendsen HJ, Hayward S (2000) Curr Opin Struct Biol 10: 165–169
18. Hsu ST, Bonvin AM (2004) Proteins 55:582–593
19. Pan Y, Ma B, Nussinov R (2005) J Mol Biol 350:514–527
20. Pan Y, Ma B, Keskin O, Nussinov R (2004) J Biol Chem 279:30523–30530
21. de Groot BL, van Aalten DMF, Scheek RM, Amadei A, Vriend G, Berendsen HJ (1997) Proteins 29:240–251
22. Barrett CP, Noble ME (2005) J Biol Chem 280:13993–14005
23. Barrett CP, Hall BA, Noble ME (2004) Acta Crystallogr D Biol Crystallogr 60:2280–2287
24. Mello LV, De Groot BL, Li S (2002) J Biol Chem 277:36678–36688
25. Vreede J, van der Horst MA, Hellingwerf KJ, Crielaard W, van Aalten DM (2003) J Biol Chem 278:18434–18439
26. Bairoch A, Apweiler R, Wu CH, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M, Martin MJ, Natale DA, O'Donovan C, Redaschi N, Yeh LS (2005) Nucl Acids Res 33:D154–D159 (Database issue)
27. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) Nucl Acids Res 28:235–242
28. Vranken WF, Budesinsky M, Fant F, Boulez K, Borremans FA (1995) FEBS Lett 374:117–121
29. Baker D, Sali A (2001) Science 294:93–96
30. Humphrey W, Dalke A, Schulten K (1996) J Mol Graph 14:33–38, 27–28
31. Sali A, Blundell TL (1993) J Mol Biol 234:779–815
32. Kabsch W, Sander C (1983) Biopolymers 22:2577–2637
33. Amadei A, Linssen ABM, Berendsen HJC (1993) Proteins 17:412–425
34. Levy R, Srinivasan A, Olson W, MsCammon J (1984) Biopolymers 23:1099–1112
35. Garcia AE (1992) Phys Rev Lett 68:2696–2699
36. Hayward S, Go N (1995) Annu Rev Phys Chem 46:223–250
37. Berendsen HJC, Van Der Spoel D, Van Drunen R (1995) Comp Phys Comm 91:43–56
38. Lindahl E, Hess B, Van Der Spoel D (2001) J Mol Mod 7: 306–317
39. Van Aalten DMF, Findlay JBC, Amadei A, Berendsen HJC (1995) Prot Eng 8:1129–1136
40. Van Aalten DMF, de Groot BL, Berendsen HJC, Findlay JBC, Amadei A (1997) J Comp Chem 18:169–181
41. Faraldo-Gómez JD, Forrest LR, Baaden M, Bond PJ, Domene C, Patargias G, Cuthbertson J, Sansom MSP (2004) Proteins 57: 783–791
42. McCammon JA, Harvey S (1987) Dynamics of proteins and nucleic acids. Cambridge University Press, Cambridge
43. Huang CC, Tang M, Zhang MY, Majeed S, Montabana E, Stanfield RL, Dimitrov DS, Korber B, Sodroski J, Wilson IA, Wyatt R, Kwong PD (2005) Science 310:1025–1028
44. Hunenberger PH, Mark AE, van Gunsteren WF (1995) J Mol Biol 252:492–503
45. Laskowski RA, MacArthur M, Moss DS, Thornton JM (1993) J Appl Crystallogr 26:283–291
46. de Groot BL, Hayward S, van Aalten DMF, Amadei A, Berendsen HJC (1998) Proteins 31:116–127