Available online at www.sciencedirect.com
ScienceDirect
Journal of Molecular Graphics and Modelling 26 (2007) 306–318
Journal of Molecular Graphics and Modelling
www.elsevier.com/locate/JMGM

Molecular motions in HIV-1 gp120 mutants reveal their preferences for different conformations

Shu-Qun Liu [a], Ci-Qun Liu [a], Yun-Xin Fu [a,b,*]

[a] *Laboratory for Conservation and Utilization of Bio-resources, Yunnan University, Kunming 650091, PR China*
[b] *Human Genetics Center, The University of Texas Health Science Center, Houston, TX 77030, USA*

Received 31 August 2006; accepted 10 December 2006
Available online 15 December 2006

Abstract

Both the crystal structures of the HIV-1 gp120 core bound by the CD4 and antigen 17b, and the SIV gp120 core pre-bound by CD4 are known. We built the homology models of gp120 with loops V3 and V4 in the CD4-complex, CD4-free and CD4-unliganded states, and models of the 375 S/W and 423 I/P mutants in the CD4-free and unliganded states, respectively. CONCOORD was utilized for generating ensembles of the seven gp120 models that were analyzed by essential dynamics analyses to identify their preferred concerted motions. The revealed large-scale concerted motions are related to either the receptor association/release or the conformational transition between different conformational states. Essential subspace overlap analyses were performed to quantitatively distinguish the preference for conformational transitions between states of the gp120 mutants and further to ascertain what kind of conformational state that the mutants prefer to adopt. Results indicate that the 375 S/W mutant, in which the tryptophan indole group is predicted to occupy the phe43 pocket in the gp120 interior, favors a conformation close to the CD4-bound state. However, the other mutant 423 I/P inclines to prevent the formation of bridging sheet and stabilize the conformation in the unliganded state. Our theoretical analyses are in agreement with experimentally determined mutation effects, and can be extended to a new approach to design or screen mutants that have effects on conformation/function of a protein.

Keywords: gp120 mutants; Conformational transition; Essential dynamics; Essential subspace overlap; Mutation effect

1. Introduction

The human immunodeficiency viruses (HIV-1 and HIV-2) and simian immunodeficiency viruses (SIV) are the etiologic agents of acquired immunodeficiency syndrome (AIDS) in their respective human and simian hosts [1–4]. Entry of the HIV-1 into target cells is initiated by the selective interaction between the viral exterior envelope glycoprotein, gp120, and the receptor on the surface of the host cell, CD4, as well as the obligatory chemokine receptors CCR5 or CXCR4 [5–7]. Accumulating biochemical and structural evidences indicate that the initial binding of the CD4 to the gp120 triggers conformational alterations in the HIV envelope that subsequently promote recognition of the coreceptor and ultimately lead to conformational changes of gp41 and membrane fusion [8–10]. This complex mechanism involves a series of structural rearrangements in which the conformational dynamics of the HIV envelope glycoproteins plays crucial role.

Both the crystal structures of the HIV-1 gp120 core [11,12] bound by the CD4 and antigen binding fragment (Fab) of the human neutralizing antibody 17b, and the SIV gp120 core [13] are known. They are conformations corresponding to post- and pre-bound by the CD4, respectively. Although both of the cores contained two domains (the inner and outer domains), the rough structural comparison revealed unexpectedly extensive conformational rearrangements upon receptor binding [13], which was in line with thermodynamics evidence suggesting that the binding of CD4 or antibodies to gp120 induced substantial structural rearrangements of the gp120 [14,15].

Although the crystallographic studies have provided two static snapshots of gp120 cores in the CD4-bound and CD4-unliganded states, it has been generally accepted that the

* Corresponding author at: Laboratory for Conservation and Utilization of Bio-resources, Yunnan University, Kunming 650091, PR China.
Tel.: +86 871 5035257; fax: +86 871 5035257.
E-mail address: shuqunliu@ynu.edu.cn (Y.-X. Fu).

1093-3263/$ – see front matter
doi:10.1016/j.jmgm.2006.12.007

conformational flexibility of HIV-1 envelope glycoproteins is important for their functions in mediating virus entry and evading humoral immune response [13–16]. The comparative analyses of molecular motions and dynamics data for gp120 in different conformational states are helpful in a full understanding of the protein function. In particular, the understanding of ranges of conformations available to the gp120 in different functional states and further distinguishing the preference for conformational transition between states are of importance for the appreciation of their functional roles and for guiding attempts at intervention. Xiang et al. [17] have identified two groups of HIV-1 gp120 mutants using the ligand-binding phenotypes to reach a conclusion that the gp120 glycoprotein can assume at least two distinct conformations. Concretely, for a mutant 375 S/W, both the enthalpy change and the entropy penalty observed when the wild-type gp120 bound to the CD4 were significantly reduced, and it was recognized better than the wild-type gp120 by the CD4 and CD4-induced (CD4i) antibodies, whereas the affinity to antibodies elicited against the CD4-binding site (CD4BS) was almost totally abolished, indicating the 375 S/W mutant favors a conformation near the CD4-bound state. Another mutant, the 423 I/P, did not bind the CD4 or CD4i antibodies, even though recognition by the CD4BS antibodies was efficient, indicating that the 423 I/P mutant favors a conformation differing from the CD4-bound state, i.e. the CD4-unliganded or ground state.

The conformational changes involved range from very subtle, local changes to global conformational changes involving motions of significant amplitude for a 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 crucial in the process of conformational transition from the CD4-unliganded state to the CD4-bound state as unusually large structural differences exist between them. 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 can be gained through analysis of computational simulated dynamics [18]. For example, previous studies on the dynamics properties of gp120 in the CD4-free and gp120–CD4-complex states using 10 ns molecular dynamics (MD) simulations [19] have demonstrated that the CD4 association reduced conformational flexibility in certain loop regions of the outer domain. Other MD studies indicated that the CD4 can merely lock the β20–β21 hairpin of the bridging sheet in gp120 but leave the β2–β3 hairpin flexible [20,21]. A shortcoming of the MD approach, however, is that the currently available computing power permits only short simulations to be run for biological macromolecules, typically of the order of tens of nanoseconds. This time scale may insufficiently explore the conformational space available to a protein, and is a few orders of magnitude smaller than that on which most biological processes take place. Thus the MD-derived ensembles are most useful in predicting both the detailed behavior and local conformational changes of a protein. An alternative approach to MD is to generate an ensemble of structures randomly without using Newton's equations of motion. The CONCOORD [22] takes this approach as a fast way to generate ensembles that explore the conformational space more fully. Previous studies [22–26] have shown that ensembles derived from CONCOORD are most useful in identifying global motions that a protein is able to perform and predicting the character of conformational changes.

We have performed the comparative modeling technique to generate the three-dimensional structural models of gp120 in three 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). We also introduced point mutation into the wild-type gp120 models to generate the 375 S/W and 423 I/P mutants in the CD4-free and unliganded states, respectively. These seven homology models were used as starting structures for CONCOORD to generate respective ensembles. One of our aims was to characterize the molecular motions of gp120 in different conformational states at global level to determine the hierarchy of tertiary structures that exists in gp120. The other aims were to investigate the variations of molecular motions caused by point mutations, and further identify the preference for conformation transition between gp120 mutants and finally, ascertain which conformational states that the 375 S/W and 423 I/P mutants prefer to adopt. Our results show that the 375 S/W mutant sampled a conformational space that is very close to the CD4-bound state, while the 423 I/P mutant preferred a conformation near the unliganded state. The consistency of our theoretical results with experimentally determined mutation effects not only compensates the mutagenesis experiments and MD studies, but also extends a new approach to design or screen mutants that have effects on conformation/function of a protein.

2. Materials and methods

2.1. Generation of the starting structures

The starting sequence of the HIV-1 HXBc2 isolate gp160 precursor (Swiss-Prot accession no. P04578) was obtained from the Swiss-Prot protein sequence database [27]. The sequence for the transmembrane glycoprotein gp41, 52 and 19 residues from the N- and C-termini and the V1/V2 loop (Gly-Ala-Gly substitutes for 67 V1/V2 loop residues) of the gp120 were removed. The final gp120 primary sequence consists of residues from 83 to 492, including sequences of the loops V3 and V4. Structural templates of the gp120 core and V3 loop were obtained from the PDB protein structure database [28], out of them PDB entries 1G9M (chain G) [11] and 2BF1 [13] were used as templates to model the CD4-bound and unliganded gp120 cores, respectively, and 1CE4 [29] for the V3 loop.

To obtain homology models of the gp120 with V3 and V4 loops in the CD4-bound conformational state, the sequence

alignment between templates of the V3 loop (1CE4) and gp120 core (1G9M, chain G) and the target sequence of the HXBc2 was first constructed by a in-house written script ‘‘alignment’’ for the MODELLER software package [30], and the alignment result was shown in Fig. 1. Second, 20 gp120 models were generated using an in-house written script ‘‘get-model’’ for MODELLER. The V4 loop coordinates were generated and optimized by a loop modeling sub-routine within the ‘‘get-model’’ with a 3D_INTERPOLATION algorithm and a thorough optimization protocol, respectively. Finally, these 20 models were clustered and optimized 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 the corresponding atom with a cluster cut off 1.5 Å. The final optimized average gp120 model is in the CD4-bound conformational state without CD4 presented, which we referred to as the CD4-free gp120 in the rest of the text. To obtain the gp120 model in complex with CD4, the CD4-free gp120 model was superimposed onto the crystal structure of the gp120–CD4 complex (1G9M) to obtain the correct orientation with respect to CD4. The D2 domain of the CD4 was removed while the CD4-free gp120 and CD4 D1 domain from the crystal complex were preserved, which we referred to as the gp120–CD4 complex in the context. The monomer gp120 within the gp120–CD4 complex was referred to as the CD4-complex gp120 in the rest of the text.

The homology model of the unliganded HIV-1 gp120 with the V3 loop was obtained as described for generating the CD4-free gp120 model except that the template is 2BF1. The final optimized average model in the CD4-unliganded conformation was referred to as the unliganded gp120 in the rest of the text.

To obtain gp120 models with mutated residues, we substituted Trp for Ser at the 375 position and Pro for Ile at the 423 position in the HXBc2 target sequence, respectively. The CD4-free and unliganded gp120 models were employed as templates to generate mutated models in the two states, respectively, and the modeling procedure resembled that of generating the wild-type gp120 models. We finally obtained four gp120 mutated models, i.e. the CD4-free 375 S/W mutant, the unliganded 375 S/W mutant, the CD4-free 423 I/P mutant, and the unliganded 423 I/P mutant.

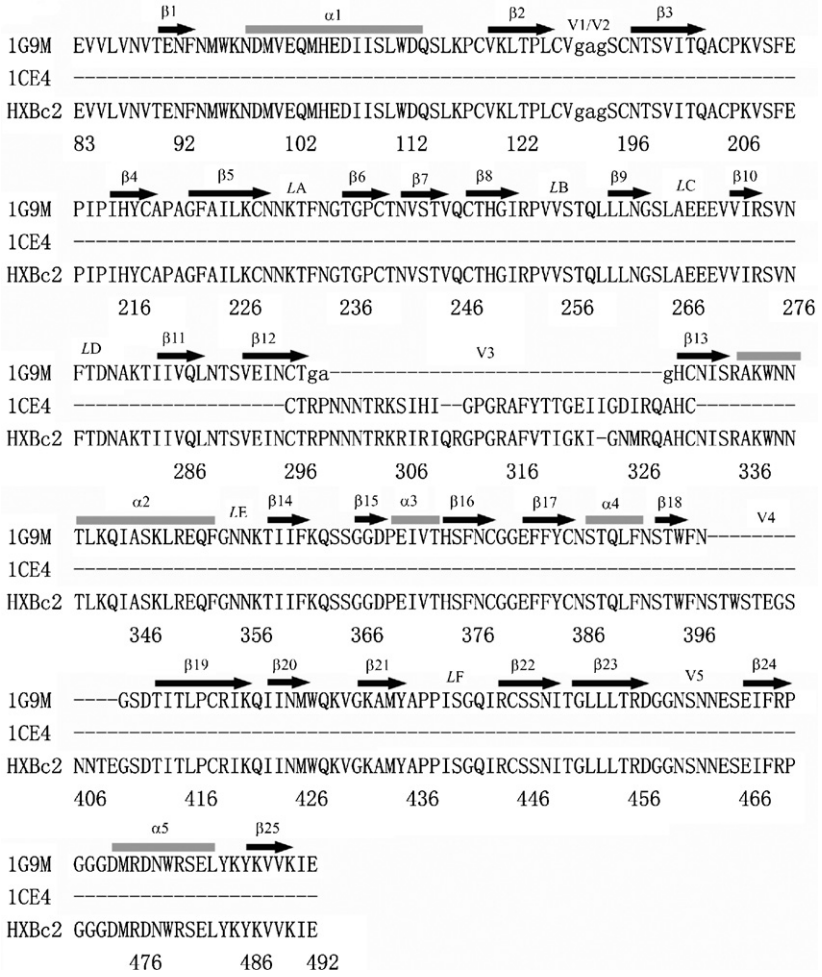

Fig. 1. Multiple sequence alignment between templates of the gp120 core (PDB entry 1G9M, Chain G) and V3 loop (PDB entry 1CE4) and target sequence of the HXBc2 isolate (Swiss-Prot accession no. P04578). The secondary structure elements are assigned according to the X-ray gp120 core with arrows for β-strands 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.

2.2. Generation of the structural ensembles

The seven gp120 homology models (gp120–CD4 complex, CD4-free, unliganded, CD4-free 375 S/W, unliganded 375 S/W, CD4-free 423 I/P, and unliganded 423 I/P) were used as starting structures for the CONCOORD program [22] to generate respective structural ensembles. CONCOORD works in two phases. In the first phase, it derives a table of upper and lower distance limits for various interatomic distances (covalent, ionic, and hydrogen bond, etc.) that define a region of conformation space that a protein can assess. These distance limits depend on distances measured in the parent structure and the strength of the interatomic interaction. For example, if a covalent bond is recognized in the parent structure, a distance limit is introduced that will reproduce this bond in all members of the derived ensemble to within a small tolerance. Similarly, a salt bridge interaction in the parent structure will be reproduced in all ensemble members, although in this case a slightly greater tolerance is permitted. In the second phase, CONCOORD generates an ensemble of structures that fulfil the 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 until all the constraints are satisfied. The generated ensemble can be thought of as a set of uncorrelated structural snapshots with the temporal order of frames scrambled, even so, in many ways it can be treated and analyzed in exactly the same way as a trajectory of snapshots produced by an MD simulation [22–26]. CONCOORD was run 7 times, once for each of the seven models generated by MODELLER, and 500 conformations were finally obtained for each of the wild-type and mutated gp120s, respectively.

2.3. Essential dynamics analysis

Essential dynamics (ED) method [31] is a powerful tool for filtering large-scale concerted motions from an ensemble of structures. It consists of a diagonalization of the covariance matrix of atomic fluctuations, resulting eigenvectors are directions in the conformational space that represent the concerted motions of which the corresponding eigenvalues define the mean square fluctuations (MSF) of the motion along these vectors. The required covariance matrices, eigenvectors, and corresponding eigenvalues of the seven CONCOORD ensembles were obtained by applying the g_covar program of the GROMACS package [32], and only the C_α atoms were included in ED analyses presented here.

The DYNAMITE program [24] was utilized for generating the ''porcupine plot'' that was able to give a graphical representation of the motion held in an eigenvector. For instance to visualize eigenvector 1, a cone is drawn for each residue starting from the C_α, which projects in the direction of the component of the first eigenvector that corresponds to that residue with the length of the cone representing the motional amplitude. Given eigenvectors of the ED decomposition, a script was generated by DYNAMITE to allow the molecular graphics program VMD [33] to automatically plot these cones onto the protein.

2.4. Essential subspace overlap analysis

If the eigenvectors obtained from the ED analysis are seen as vectors that span a complex $3N \times 3N$ (with N = the number of atoms in the protein) dimensional space, then a few ''essential'' eigenvectors with the largest eigenvalues span a subspace, the essential subspace, and all large concerted motions take place in this subspace. Previous studies have shown that these large concerted motions usually play crucial role in the function of proteins take place [25,26,34,35], and furthermore, the essential subspace overlap between different simulated systems can provide a valid method to assess their dynamical [22,34,36] or functional similarity [26]. To do so, a common approach is to select a subset of eigenvectors for each ensemble, and to calculate their cumulative mean square inner product (CMSIP) using the expression [37]:

$$\psi = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}(V_i^{\mathrm{A}} \cdot V_j^{\mathrm{B}})^2$$

where the V_i^{A} and V_j^{B} are the ith and jth eigenvectors of two different sets of ensemble, respectively, and the $(V_i^{\mathrm{A}} \cdot V_j^{\mathrm{B}})^2$ is the square inner product. The subspace overlap ψ thus ranges from 0, when the eigenvector subsets are completely dissimilar, to 1 (or 100%) when they are identical. The number of eigenvectors used, n, is typically chosen so as to represent a significant proportion of fluctuations in the simulation.

We performed the essential subspace overlap analyses between the CONCOORD ensembles of the seven gp120 models to quantify the extent to which that the gp120 in different conformational states explore the same conformational space, and furthermore, quantitatively distinguish the preference for conformational transition between mutants in different conformational states.

3. Results

3.1. Descriptions of gp120 models in the CD4-complex, CD4-free and unliganded states

A ribbon diagram of the CD4-free gp120 was shown in Fig. 2(a). The structure closely resembles its template, the gp120 core from 1G9M [11] with the exception that the V3 and V4 has been modeled on it, and the root mean square deviation (RMSD) between the shared backbone atoms of the model and template is 0.36 Å. The CD4-free gp120 comprises two major domains according its structural organization, i.e. the inner and outer domains with some excursions emanating from them. The inner domain includes the N- and C-termini, a two-helix, a small two-stranded bundle, and a six-stranded β-sandwich at its termini-proximal end and a projection at its 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. The V3 loop connecting the β12 and β13 lies beneath

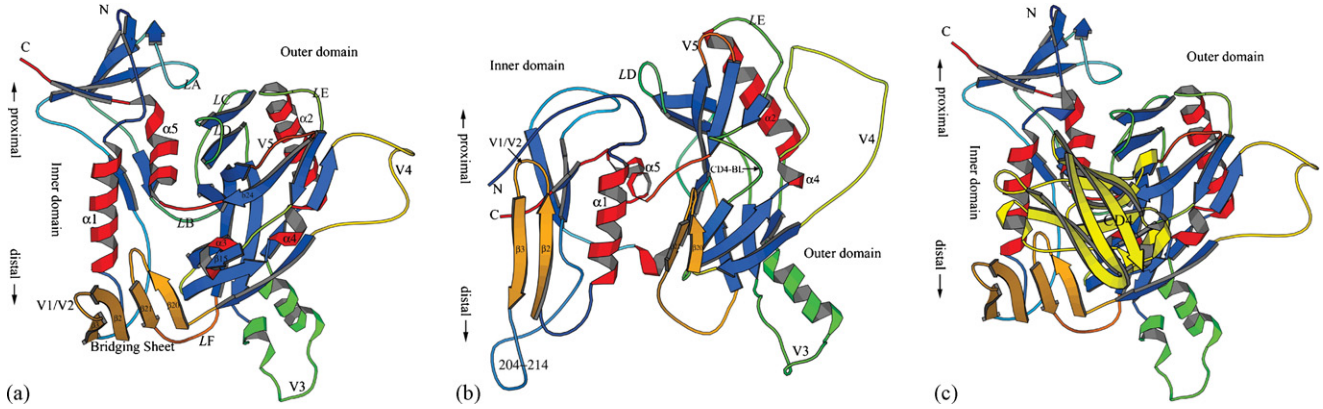

Fig. 2. Structures of the gp120 homology models. (a) Ribbon diagram of the CD4-free gp120, with α-helices in red, β-strands in blue, bridging sheet in orange, and V3 loop in green. (b) Ribbon diagram of the unliganded gp120, with the structure elements in the same color as in (a). (c) Ribbon diagram of the gp120-CD4 complex, with the structure elements of gp120 in the same color as in (a) and (b), and the CD4 molecular in yellow.

the distal end of the outer domain with the distance between the highly conserved Pro-Gly of the V3 tip and the V3 base about 30 Å. Although certain conformational differences in the V3 loops exist between our gp120 model and the most recently determined structure of a V3-containing gp120 core [38], 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 coiled conformation. In the CD4-free gp120 models, the V1/V2 stem and β20–β21 constitute the antiparallel, four-stranded bridging sheet that stands below the distal ends of both the inner and outer domains.

Fig. 2(b) shows the unliganded HIV-1 HXBc2 gp120 model, which resembles its template, the unliganded SIV gp120 core [13] with the RMSD of their shared backbone atoms 0.82 Å. The unliganded gp120 model also has two major domains with the inner domain consisting of the N- and C-termini, a three-stranded β-sheet at its termini-proximal end, V1/V2 stem (β2–β3 ribbon), α-helix 1 (α1), and a short α-helix (α5) at the inner/outer domain junction. The unresolved connecting segment (residues 218–228) between the β3 and β5 in the SIV gp120 core corresponds to the residues 204–214 in the unliganded HIV-1 gp120, which was generated and optimized using the MODELLER loop-modeling subroutine and presented a disordered loop conformation extending away from the V1/V2 stem base of the inner domain. The outer domain has, with some local exceptions, globally the similar 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 between the CD4-free and unliganded gp120 in the outer domain are: (i) the lengths in loops, for instance, the V4, V5, LE are slightly longer in the unliganded form; (ii) the linker 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. Although the two β-hairpins, the β2–β3 and β20–β21 that can form the bridging sheet is ordered, a space of 22–29 Å intervenes between them, thus the bridging sheet presented in the CD4-free gp120 is absent in the unliganded gp120.

The gp120–CD4 complex was shown in Fig. 3(c), the structural organization of the CD4–complex gp120 is identical to that of the CD4-free gp120 except for the D1 domain of the CD4 being docked onto it.

The CD4-free and unliganded gp120 models with mutated residues 375 S/W and 423 I/P are similar to the respective wild-type homology models (data not shown).

In the CD4-free gp120 model, the receptor CD4 binding sites are composed of the β20–β21 ribbon, V1/V2 stem, loops LD, LE, V5, β15-α3 excursion, and linker of the β24-α5 that generate an unusually large CD4 binding cavity (Fig. 2(a)). The CD4 Phe 43 binding pocket is located at the center of the 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 Phe 43 pocket are primarily from the β20–β21 ribbon and its connection, β15-α3-β16, β24-α5 linker, and β8–β9 linker (LB). The hydrophilic residue Ser 375 located at the N-terminus of the β16 is surrounded by hydrophobic residues in the Phe 43 pocket, therefore replacement of the Ser 375 with Trp will promote stabilizing the hydrophobic interaction while filling the Phe 43 pocket more effectively with increased side chain volume. In the unliganded form, the α1, α5, β20–β21 ribbon, and the CD4-BL create a long, narrow channel at the intersection surfaces of the inner and outer domains (Fig. 2(b)), however, the CD4 cavity and CD4 Phe 43 pocket are absent in the unliganded state. The Ile 423 lies at the N-terminus of the β20, replacement of this Ile with a Pro will place conformational strain on β20 and hinder formation of bridging sheet.

3.2. Essential dynamics analyses

The CONCOORD conformation ensembles generated from the seven gp120 models were subject to ED analyses, and the eigenvalues corresponding to the first 10 eigenvectors and the traces (total MSF) of the covariance matrices after diagonalizations were shown in Table 1. In all cases only a few eigenvectors

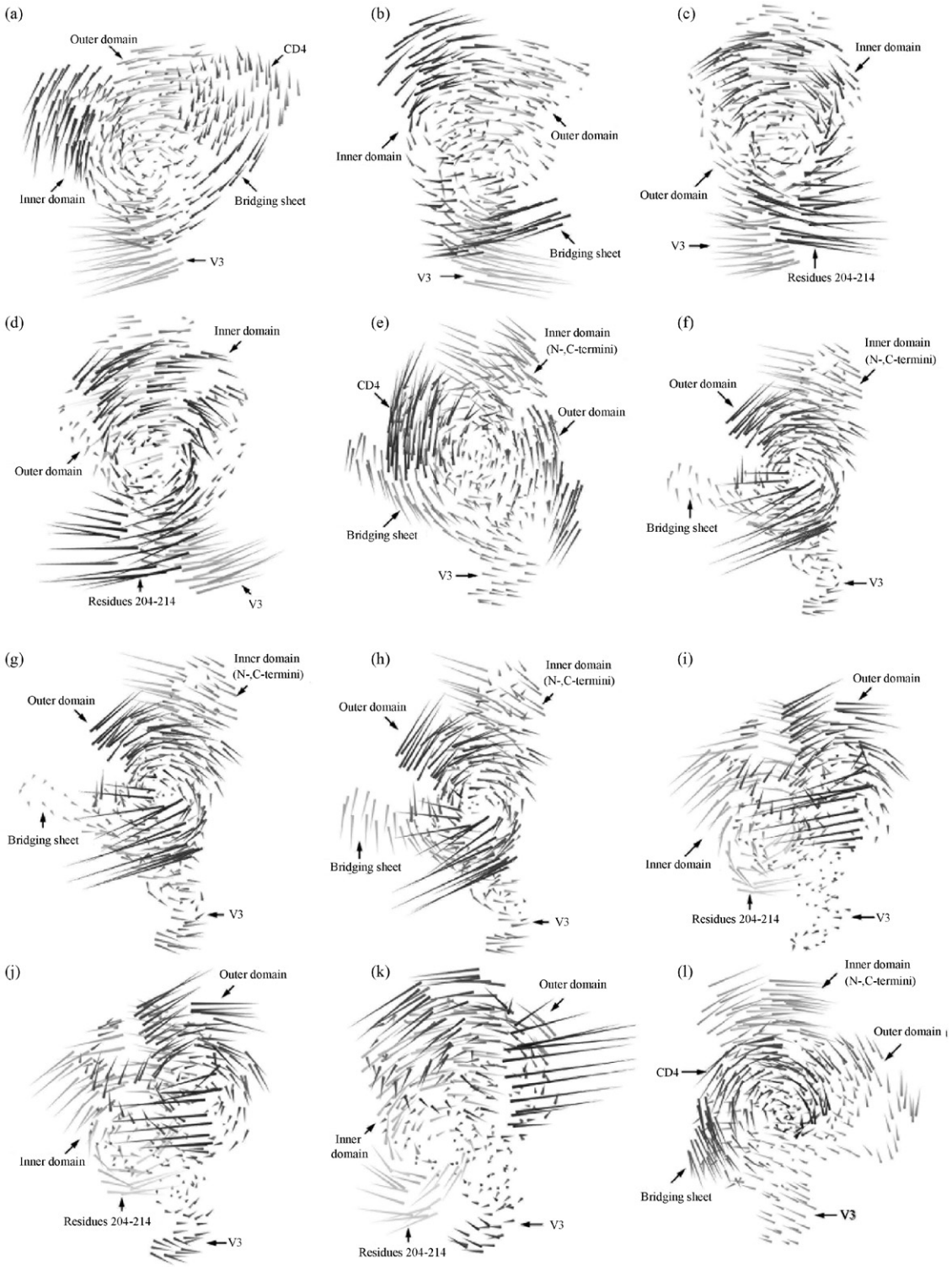

Fig. 3. Porcupine plots of the principal modes of conformational variability calculated from the CONCOORD ensembles of the seven gp120 models. The most significant motion described by eigenvector 1 for (a) the gp120-CD4 complex, (b) the wild-type CD4-free gp120, (c) the wild-type unliganded gp120, and (d) the unliganded 423 I/P mutant. The most significant motions of the CD4-free 375 S/W and 423 I/P mutants are identical to that of the wild-type CD4-free gp120 shown in (b). The most significant motion of the unliganded 375 S/W mutant is identical to that of the wild-type unliganded gp120 shown in (c). These views are all from the inner to outer domains. The second ranked motion described by eigenvector 2 for (e) the gp120-CD4 complex, (f) the wild-type CD4-free gp120, (g) the CD4-free 375 S/W mutant, (h) the CD4-free 423 I/P mutant, (i) the wild-type unliganded gp120, (j) the unliganded 375 S/W mutant, and (k) the unliganded 423 I/P mutant. These views are all from the outer domain to inner domain. The third ranked motion described by eigenvector 3 for (l) the gp120-CD4 complex, (m) the wild-type CD4-free gp120, (n) the CD4-free 375 S/W mutant, and (o) the wild-type unliganded gp120. The third ranked motion of the CD4-free 423 I/P mutant is identical to that of the wild-type CD4-free gp120 shown in (m). The third ranked motions of the unliganded 375 S/W and 423 I/P mutants are identical to that of the wild-type unliganded gp120 shown in (o). The views for (l), (m), and (n) are looking down from the CD4 molecule or CD4 binding cavity. The view for (o) is from the inner to outer domains.

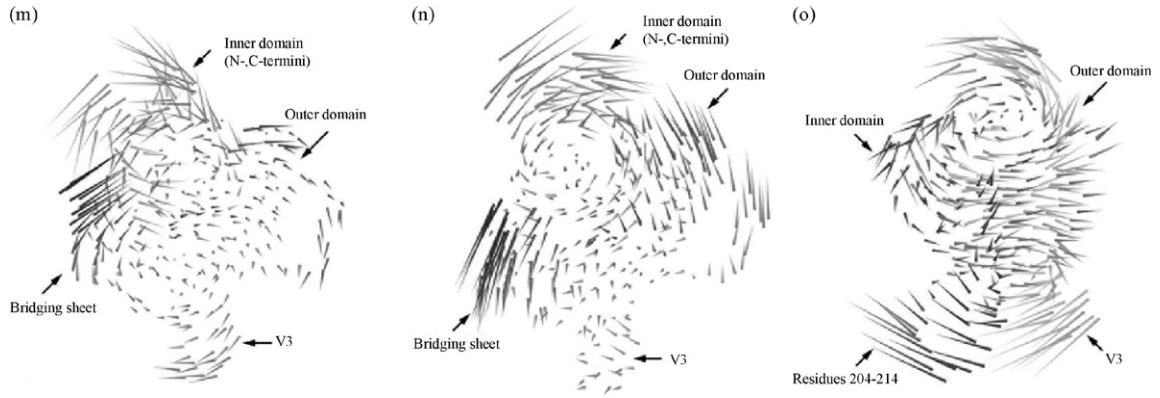

Fig. 3. (*Continued*).

were found with significant eigenvalues, i.e. the contributions of the first four eigenvectors to the total MSF are 65.7%, 70.0%, 70.5%, 70.4%, 72.8%, 68.8%, and 71.3%, and of the first 10 eigenvectors to the total MSF are 84.3%, 86.5%, 84.3%, 86.7%, 85.1%, 86.3%, and 85.2% for the ensembles of the wild-type CD4-complex, CD4-free, and unliganded gp120s as well as the CD4-free 375 S/W, unliganded 375 S/W, CD4-free 423 I/P, and unliganded 423 I/P mutants, respectively, indicating that the first 10 eigenvectors, especially the first four eigenvectors, emerge with appreciable freedom.

Fig. 3 shows, in porcupine representation, the large-scale concerted motions of the seven gp120 ensembles. The most significant motions of the wild-type and mutated models in different conformational 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 seven ensembles will result in different functional effects. For the gp120–CD4 complex (Fig. 3(a)), the gp120 inner domain and bridging sheet together with the 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 the twist will grip the CD4 molecule tightly because the bridging sheet and the components (in particular the LD, V5, and LE) lied at the proximal end of the outer domain move towards the CD4. For CD4-free models of the wild-type gp120 (Fig. 3(b)) and 375 S/W and 423 I/P mutants (data not shown, the porcupine plots of the mutants are identical to Fig. 3(b)), 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–complex gp120. This mode will enlarge the CD4 cavity because the bridging sheet and the components lied at the proximal end of the outer domain move away from the cavity. For unliganded models of the wild-type gp120 (Fig. 3(c)) and 375 S/W mutant (data not shown, the porcupine plot is identical to Fig. 3(c)), the twisting motion is similar to that of the CD4-complex gp120 despite of the large structural differences between them, i.e. the anticlockwise vortex formed by the inner domain rotates with respect to the clockwise vortex formed by the outer domain. Although this twisting mode in the unliganded gp120 does not directly open or close the long, narrow channel, the shared rotating directions by the two vortices in the CD4-complex and unliganded gp120s may impart the unliganded wild-type and 375 S/W gp120s the potency to translate their conformations into the CD4-complex state, or to sample conformations that can favourably associate with the CD4. However, for the unliganded 423 I/P mutant (Fig. 3(d)), the two vortices formed

Table 1
Eigenvalues corresponding to the first 10 eigenvectors and the traces of the covariance matrices obtained from the CONCOORD ensembles of the wild-type CD4-complex, CD4-free, unliganded gp120s and the CD4-free 375 S/W, unliganded 375 S/W, CD4-free 423 I/P, and unliganded 423 I/P mutants

Eigenvectors	Wild-type			375 S/W		423 I/P	
	CD4-complex	CD4-free	Unliganded	CD4-free	Unliganded	CD4-free	Unliganded
1	7.17	10.87	12.47	11.04	14.76	9.78	13.79
2	2.82	3.98	3.28	3.81	3.61	3.66	3.49
3	1.66	1.87	1.85	2.03	2.15	2.04	2.12
4	1.60	1.721	1.32	1.80	1.63	1.72	1.40
5	1.11	1.40	0.90	1.24	0.88	1.28	0.99
6	0.82	0.90	0.74	0.89	0.81	0.90	0.78
7	0.67	0.68	0.64	0.72	0.72	0.66	0.70
8	0.53	0.57	0.57	0.59	0.53	0.63	0.61
9	0.32	0.47	0.45	0.49	0.48	0.53	0.55
10	0.29	0.34	0.39	0.36	0.40	0.35	0.45
Trace	20.19	26.35	26.82	26.51	30.41	25.00	29.18

by the inner and outer domains rotate in opposite directions compared to those of the unliganded wild-type gp120 and 375 S/W mutant, implying the 423 I/P mutant inclines to preserve its unliganded conformation that is relatively hard to associate with the CD4.

All the second eigenvectors of the seven ensembles are dominated by a rotation/twisting motion of the outer domain with respect to the inner domain, although the bridging sheet and V3 loop in the CD4-complex and CD4-free (including the wild-type and mutants) gp120s rotate in concert with the inner domain. For the gp120–CD4 complex (Fig. 3(e)), a large part of the outer domain (except for the V3 loop) together with the CD4 forms a clockwise vortex, which rotates relative to an anticlockwise vortex formed by the inner domain, the bridging sheet, and the V3 loop. For CD4-free models of the wild-type gp120 (Fig. 3(f)), 375 S/W (Fig. 3(g)) and 423 I/P (Fig. 3(h)) mutants, the clockwise vortices formed by a large part of the outer domain (except for the V3 loop) rotate relative to the anticlockwise vortices formed by the inner domain, bridging sheet, and V3 loops in the same direction as that in the CD4-complex gp120. This mode will enlarge the CD4-binding cavity and may be related to the CD4 release for reasons that the bridging sheet and components located at the proximal end of the outer domain move away from the CD4 cavity. For unliganded models of the wild-type gp120 (Fig. 3(i)) and 375 S/W mutant (Fig. 3(j)), the rotating directions of the two vortices formed by the inner and outer domains resemble those of the CD4-complex (Fig. 3e) and CD4-free gp120s (Fig. 3(f)–(h)), which may be related to the conformational transition from the unliganded state to the complex state. However, this twisting mode will slightly close the long, narrow channel. For the unliganded 423 I/P mutant (Fig. 3(k)), the two vortices formed by the inner and outer domains rotate in opposite directions compared to those of the unliganded wild-type gp120 (Fig. 3(i)) and 375 S/W mutant (Fig. 3(j)), which will, on the one hand, open the long, narrow channel, and on the other hand hinder the unliganded 423 I/P mutant from transition to the CD4-complex state because the V1/V2 stem moves away from the outer domain that prevent the progress of formation of the bridging sheet, whereas in the unliganded 375 S/W mutant and wild-type gp120, the bridging sheet inclines to form since the V1/V2 stem moves towards the outer domain. Of particular interest is that the bridging sheet emerges respectively with the largest and smallest motional amplitude in the 423 I/P (Fig. 3(h)) and 375 S/W mutant (Fig. 3(g)) among the three CD4-free models (Fig. 3(f)–(h)), we also note that in all the three CD4-free models the bridging sheets move away from the CD4 cavity, therefore the highest mobility of the bridging sheet in the CD4-free 423 I/P mutant will promote conformation transition towards the unliganded state, while the relative stabilization of the bridging sheet in the CD4-free 375 S/W mutant will prevent conformational transition towards the unliganded state to some extent. In addition, the reduced flexibility of the bridging sheet caused by the 375 S/W mutation might also be responsible for the reduced entropic change for the CD4 binding. We also note that the V3 loop and LF in the CD4-complex gp120 (Fig. 3(e)) emerge with the largest motional amplitude among the three wild-type gp120 models (Fig. 3(e), (f), and (i)), we speculate that the high mobility of the LF and V3 is caused by CD4's association and would be of great benefit to promoting association of gp120 with the coreceptor CCR5/CXCR4 or CD4i antibodies. Interestingly, among the four gp120 mutants (Fig. 3(g), (h), (j), and (k)), the highest mobility of the LF and V3 occurs in the CD4-free 375 S/W (Fig. 3(g)) and unliganded 375 S/W (Fig. 3(j)) models, indicating that the 375 S/W mutant is more readily recognized by coreceptors or CD4i antibodies than the 427 I/P mutant.

The third eigenvector of the gp120–CD4 complex (Fig. 3(l)) describes the rotation of an anticlockwise vortex formed by the CD4, bridging sheet, and a large part of the outer domain (except for the V3 loop) relative to another clockwise vortex formed by the inner domain and V3 loop. This motion will close the narrow part of the CD4 cavity that is located at the interfaces (composed of the α5, LC, LD, β9, β10, and β11, see Fig. 2(a)) of the proximal ends of the inner and outer domains while open the CD4 Phe 43 pocket that is located at the intersection of the outer domain, inner domain, and bridging sheet, and may be related to the orientation of the CD4 and the insertion of the CD4 Phe 43 into its pocket. The third eigenvectors of the CD4-free models of the wild-type gp120 (Fig. 3(m)) and 423 I/P mutant (data not shown, the porcupine plot is identical to the Fig. 3(m)) are dominated by a relatively larger clockwise vortex formed by the bridging sheet and the entire outer domain that rotates around an axis running through the Phe 43 pocket, while the inner domain N-, C-termini behave a flexing in the opposite direction relative to this vortex. This mode leads to reversed effects on the narrow part of the CD4 cavity and the Phe 43 pocket, namely enlarges the former and reduces the latter, and may also be related to the conformational transition from the CD4-free form to the unliganded form because the moving directions of the N-, C-termini and the vortex are reversed compared to those of the CD4-complex gp120. The rotating directions of the two vortices of the CD4-free 375 S/W mutant (Fig. 3(n)) resemble those of the CD4-complex gp120, indicating the CD4-free 375 S/W mutant tends to retain its CD-bound conformation. The third eigenvector of the unliganded wild-type gp120 (Fig. 3(o)) is dominated by a single anticlockwise vortex formed by the large parts of the inner and outer domains, the other regions such as the residues 204–214 in the inner domain and the V3 loop in the outer domain behave flexing motions in the opposite direction relative to the single vortex. This kind of flexing will influence reorientations of the V1/V2 stem, the V3 loop, and the connecting segment (residues 204–214) between the V1/V2 stem and the three-stranded bundle at the termini-proximal end of the inner domain, and likely be related to the approach of the V1/V2 stem to the β20–β21 ribbon that is able to drive the formation of the bridging sheet. The third eigenvectors of the unliganded gp120 mutants (375 S/W and 423 I/P) describe a motional mode that is identical to that of the unliganded wild-type gp120 (data not shown).

The fourth eigenvectors of all the seven gp120 models describe the elongations or shortnesses of the inner domain and the outer domain along respective major axes (data not shown).

These motions will lead to the thinnings or condensations of the two domains that may contribute to conformational transition between the three conformational states.

3.3. Essential subspace overlap and conformational transition

Comparisons of the motions of the seven gp120 ensembles were carried out by calculation of the essential subspace overlap between eigenvectors derived from the diagonalizations of covariance matrices because the essential subspace is of particularly interest. As described above, for all of the simulation systems studied, the first 4 and 10 eigenvectors occupy more than 65% and 80% of the total MSF, respectively, indicating the subspaces spanned by the first 10 eigenvectors, especially by the first 4 eigenvectors, will cover the majority of modes of the motion (Table 1), namely the ''essential'' motions. Therefore, the cumulative mean square inner products between the first 4 eigenvectors of one conformational state and the first 10 eigenvectors of another conformational state were calculated (Table 2), which provides a strictly quantitative measure for evaluating the preference of conformational transition between states. Specially, the essential subspace overlaps are able to evaluate the variations of molecular motions caused by point mutations, and furthermore, ascertain which conformational states that the 375 S/W and 423 I/P mutants prefer to adopt.

For the wild-type CD4-complex and CD4-free gp120 ensembles, Table 2 shows that the fist 4 eigenvectors of CD4-complex ensemble are to a large extent (83.8%) presented in the subspace spanned by the first 10 eigenvectors of the CD4-free gp120 ensemble. Under the reversed condition, the first 4 eigenvectors of the CD4-free gp120 ensemble are reproduced 85% by the first 10 eigenvectors of the CD4-complex gp120 ensemble. The large overlaps demonstrate that all the essential motions occurring in the CD4-complex gp120 also occur in the CD4-free gp120, and vice visa. For the wild-type unliganded and CD4-complex gp120 ensembles, the first 10 eigenvectors of the unliganded gp120 ensemble can reproduce 26.9% of the essential subspace spanned by the first 4 eigenvectors of the CD4-complex ensemble, and the value for the reversed overlap is 22.0%. The small overlaps indicate marked differences exist between the large-scale motional freedoms of the CD4-complex and unliganded gp120s, and the probable reason is that the large conformational differences between them (see Fig. 2). However, all the eigenvectors of the unliganded gp120 ensemble can reproduce ~90% of the first four eigenvectors of the CD4-complex ensemble, and the reversed overlap is also ~90%, implying that the unliganded gp120 has the potential to translate its conformation into the CD4-complex state, and vice versa. Furthermore, the unliganded gp120 has a larger potential to translate its conformation into the CD4-complex state because of the relatively larger essential subspace overlap (26.9%), while the potential for the CD-complex gp120 transition to the unliganded state is relatively weak because of the smaller essential subspace overlap (22.0%). For the wild-type CD4-free and unliganded gp120s, the potential for the CD4-free gp120 transition to the unliganded gp120 (25.9%) is larger than that for the unliganded gp120 transition to the CD4-free gp120 (18.0%).

The first 10 eigenvectors of the unliganded ensembles of the wild-type gp120, 375 S/W and 423 I/P mutants can reproduce 18.0%, 21.2%, and 16.1% of the essential subspace spanned by the first four eigenvectors of the wild-type CD4-free gp120 ensemble, respectively, in addition, essential subspace overlaps of unliganded ensembles of the wild-type gp120, 375 S/W and 423 I/P mutants against the CD4-complex gp120 ensemble are 26.9%, 28.6%, and 25.7%, respectively, indicating the unliganded gp120s have a greater potential to translate their conformations into the CD4-complex state than into the CD4-free state. The largest overlaps of the unliganded 375 S/W ensemble against the wild-type CD4-free (21.2%) and CD4-complex ensembles (28.6%) imply the unliganded 375 S/W mutant has the largest conformational transition potential towards the CD4-bound state among the three unliganded gp120 models. On the contrary, the smallest overlaps of the unliganded 423 I/P ensemble against the wild-type CD4-free (16.1%) and CD4-complex ensembles (25.7%) imply the unliganded 423 I/P mutant has the smallest transition potential towards the CD4-bound state but prefers maintaining its

Table 2
Cumulative mean square inner products between the first 4 eigenvectors of one gp120 conformational ensemble and the first 10 eigenvectors of another gp120 ensemble

Eigenvector 1–4	Eigenvector 1–10						
	Wild-type			375 S/W		423 I/P	
	CD4-free	Unliganded	CD4-complex	CD4-free	Unliganded	CD4-free	Unliganded
CD4-free[a]	1.00	0.180	0.850	0.877	0.212	0.846	0.161
Unliganded[a]	0.259	1.00	0.220	0.231	0.797	0.274	0.929
CD4-complex[a]	0.838	0.269	1.00	0.851	0.286	0.722	0.257
CD4-free 375 S/W[b]	0.866	0.179	0.840	1.00	0.193	NC[c]	NC[c]
Unliganded 375 S/W[b]	0.235	0.787	0.321	0.229	1.00	NC[c]	NC[c]
CD4-free 423 I/P[b]	0.854	0.202	0.709	NC[c]	NC[c]	1.00	0.147
Unliganded 423 I/P[b]	0.258	0.930	0.297	NC[c]	NC[c]	0.240	1.00

[a] Wild-type gp120.
[b] Mutant gp120.
[c] The overlap values were not calculated.

unliganded state. For all the CD4-free gp120 ensembles, the overlap of the 375 S/W mutant against the wild-type gp120 (87.7%) is larger than that of the 423 I/P mutant against the wild-type gp120 (84.6%), in addition, among the overlaps of all the CD4-free ensembles against the CD4-complex gp120, the largest overlap occurs between the CD4-free 375 S/W and CD4-complex gp120 ensembles (85.1%), while the smallest overlap occurs between the CD4-free 423 I/P and the CD4-complex gp120 ensembles (72.2%). These results indicate that the CD4-free 375 S/W mutant has the largest potential to retain the CD4-bound state among the CD4-free wild-type gp120 and mutants, while the CD4-free 423 I/P mutant has the smallest potential to do so. For the unliganded mutants, the first 10 eigenvectors of the 375 S/W and 423 I/P ensembles can reproduce 79.7% and 92.9% of the first 4 eigenvectors of the unliganded wild-type ensemble, respectively, indicating that the 423 I/P mutant inclines to maintain the unliganded conformation much more than the 375 S/W mutant does. Among the overlaps of the essential subspace between all the CD4-free/CD4-complex ensembles and the unliganded wild-type gp120 ensemble, the largest value (27.4%) occurs between the CD4-free 423 I/P mutant and the unliganded wild-type gp120, indicating the CD4-free 423 I/P has the largest conformational transition potential towards the unliganded state, the smaller overlapping value (23.1%) between the CD4-free 375 S/W mutant and unliganded wild-type gp120 indicates the relatively weak potential for the CD4-free 375 S/W to transition to the unliganded state, and furthermore, the smallest overlap (22.0%) of the CD4-complex gp120 ensemble against the unliganded wild-type gp120 ensemble implies that the CD4-complex gp120 has the smallest potential to translate its conformation into the unliganded state, indicating the CD4's association will stabilize its conformation in the CD4-bound state. The conformational transition preferences between states inferred from the essential subspace overlaps are in agreement with above qualitative analyses of the molecular motions and porcupine plots.

4. Discussions

The prerequisite of the analysis of the gp120 molecular motion and conformational transition is to obtain in depth knowledge of the structures in both the CD4-bound (liganded) and CD4-unliganded states. The 100% sequence identity between the HXBc2 target sequence and its template 1G9M results in an accurate core structural model of the liganded gp120 with approximately accurate V3 and V4 loops being modeled on this core as more than 98% residues in this model are located in the favoured regions of the Ramachandran plot obtained from the PROCHECK program [39]. The direct experimental data of the unliganded HIV-1 gp120 structure is not available at this state, however, given the high sequence similarity (over 35% sequence identity and 70% sequence similarity) between SIV and HIV, the unliganded HIV-1 gp120 is likely to assume a similar conformation because it was noted [40] that it is able to generate medium-accuracy unliganded HIV-1 gp120 models where ~90% of the main-chain atoms can be modeled with 1.5 Å RMS error under such a sequence similarity. In our unliganded HIV-1 gp120 model with the modeled V3 loop, there are more than 90% residues that have appropriate dihedral angle distributions. Furthermore, the relatively high overall sequence conservation between SIV and HIV, their common receptor and coreceptors, and as described in the literature [13], the crystal core of the SIV gp120 is thus a good representation of the conformation in the CD4-unliganded state, therefore most of the differences between our HIV-1 homology models of the unliganded gp120 and CD4-free/CD4-complex gp120 should reflect the conformational changes induced by the receptor binding. The amino acid point mutations, 375 S/W and 423 I/P were introduced into the wild-type CD4-free and unliganded gp120s respectively to generate the models of gp120 mutants, thus the seven gp120 models in different conformational states are suitable for exploring molecular motions and conformational transitions of the gp120.

To study the mobility and flexibility of a protein, simulation techniques such as the MD or CONCOORD are required to generate ensembles of a structure in which each frame contains an accessible conformation. Although the MD structures are in principle free to access all physically obtainable conformations, the relatively short timescales that this approach can currently access; the incremental search protocol that is used to obtain individual members of the ensemble; the Brownian and viscous effects of the solvent environment, may limit the full exploration of the conformational space available to a protein [23]. On the contrary, CONCOORD structures are derived from distance 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 [22]. Thus, the CONCOORD ensembles are most useful in identifying globally large-scale motions of a protein and can also be extrapolated to predict the character of conformational changes and further predict the preference for conformational transition between different conformational states of a protein.

The ED analyses of the seven CONCOORD ensembles reveal that the first four eigenvectors contribute to a significant proportion of the total fluctuation. It is worth pointing out, although the ED technique yields individual concerted motional modes, all the modes interplay with each other in a complex manner, and therefore, not necessarily all individual modes will correspond to a specific functional task. In addition, the large-scale concerted motions described by the first four eigenvectors of the seven gp120 ensembles do not represent explicit rotation/twisting or breathing (open/close) modes between the two domains as have been observed in the T4 lysozyme [35], the hyaluronate lyase [25], and the cyclin dependent kinase 2 [23]. Although gp120 structures in the CD4-bound and unliganded states can be divided into two major domains based on their structural organizations [11,13], ED analyses demonstrate that some structural components such as the bridging sheet, V1/V2 stem, and V3 loop can move in concert with either the inner domain or the outer domain,

presenting complicated combinations of variable motional modes such as rotation/twisting, opening/closing, or closing/flexing of one mobile unit (or vortex) with respect to the other mobile unit (or vortex). We suggest that these perplexing concerted motions are related to the dynamics mechanism of gp120-receptor association/release, and also provide one probable explanation for the failure of gp120 to induce potent neutralizing antibodies against the CD4 and coreceptor binding sites. In particular, the variations of motional directions and motional amplitudes of certain mobile units (or vortices) among the seven simulation systems would be related to the conformational transition from one state to the other or render some states the potencies to retain their current conformations.

It has been generally accepted that the HIV-1 gp120 glycoprotein is conformationally flexible, and the CD4's association leads the gp120 to adopt the CD4-bound conformation that is recognized by the second receptor, CCR5 or CXCR4. In some cases, however, the HIV-1 gp120s exhibit the ability to bind chemokine receptors in the absence of CD4 [41], and the SIV also often exhibit some degree of the CD4 independent infection [17], implying that the unliganded gp120 can fluctuate spontaneously towards the CD4-bound state. Stabilization of the gp120 in its CD4-bound conformation has been studied by the mutagenesis experiment [17] and MD computational simulation [20]. As has been argued by Kwong et al. [11], the hydrophobic core that was located at the bottom of the CD4 Phe 43 pocket held together the inner domain, outer domain, and the bridging sheet; it might be sensitive to a change in the environment because of the unique organization among the interaction residues, therefore the 375 S/W mutation was designed to fill the partially occupied CD4 binding pocket so that the stability or the conformation of the gp120 would be altered. The mutagenesis experiment [17] showed that the 375 S/W mutant was bound more efficiently by the CD4, CCR5, or CD4i antibodies compared to the wild-type gp120, whereas both the enthalpy change and the entropy penalty were significantly reduced for the CD4 binding, revealing that this mutant inclines to adopt a conformation that is close to the CD4-bound state. The MD study [20] explained the surprising experimental behavior of the 375 S/W mutant, i.e. the 375 S/W mutation site was surrounded by several hydrophobic/aromatic residues, the mutated Trp will make its indole group occupy the Phe 43 pocket and strengthen the hydrophobic interactions in its vicinity, which will result in the reduction of the flexibility of residues near the mutation site and be responsible for the partial reduced entropic change, and finally, stabilize the conformation of the 375 S/W mutant near the CD4-bound state. The strategy attempting to destabilize the CD4-bound conformation or stabilize the CD4-unliganded conformation was to introduce point mutation into the V1/V2 stem or β20–β21 ribbon to hinder the formation of bridging sheet. The mutagenesis experiment [17] revealed that the 423 I/P mutant, in which the gp120 bridging sheet was disrupted, did not bind the CD4, CCR5, or CD4i antibodies but bound the CD4BS antibodies efficiently, indicating this mutant inclined to maintain its CD4-unliganded conformation state. According to this, we also attempted to evaluate the point mutation effects on the gp120 structure/function using the theoretical calculation approach based on the ED technique and the conformational space overlap analysis.

Although analyses of the molecular motions of the seven simulation systems provided the qualitative description of the probable preference for conformational transition caused by point mutations, the quantitative insights into the preference for conformational transition were obtained by comprehensively comparative analyses of the essential subspace overlaps between ensembles of the wild-type gp120 and mutants. For ensembles of the wild-type gp120 and mutants in the same starting conformational state, we found the essential subspace overlaps between them were large (usually more than 70%), while for ensembles of the wild-type gp120 and mutants in the different starting conformational states, the overlaps were small (usually less than 30%). The large structural differences between the CD4-bound (CD4-free/CD4-complex) and unliganded forms are responsible for the small values of the essential subspace overlaps, indicating the apparent differences exist between the motional freedoms of the CD4-bound and CD4-unliganded states. As described above, CONCOORD generates structures by the random search method that searches for solution in a predefined coordinate space with all the interactions implemented in the form of distance constraints [26], therefore only the short-range interactions within the protein make a serious contribution. We infer that the overestimate of the non-covalent bond or short-range interaction in CONCOORD is also responsible for the small essential subspace overlap between ensembles derived from different starting conformations. Even so, here we mainly focus on the variations of the values for the similar essential subspace overlaps, i.e. the overlaps of all ensembles derived from the same starting conformational state against another ensemble from a different starting sate. Although the differences between the similar overlapping values are small, they are significant for reasons that CONCOORD hardly suffers from a sampling problem when at least a few hundred structures are generated [22]. Thus the essential subspace overlapping value is a sensitive indicator for evaluating the dynamics similarity and conformational transition preference for a protein in different starting conformational states. For example, for the similar overlaps of all unliganded ensembles of the wild-type gp120, 375 S/W, and 423 I/P mutants against the CD4-complex gp120 ensemble, the overlapping values are 26.9%, 28.6%, and 25.7%, respectively, although the differences between these values are small, the largest overlap occurs between the unliganded 375 S/W mutant and the wild-type CD4-complex gp120, we can inferred from which that the unliganded 375 S/W mutant has the largest conformational transition potential towards the CD4-bound state. According to this overlapping principle, we also inferred that the wild-type CD4-free gp120 has the larger potential to translate its conformation into the unliganded state than the CD4-complex gp120 does, which is in agreement with the MD data [20,21] and chemical intuition that the removal of CD4 from the gp120–CD4 complex would destabilize the CD4-bound conformation, resulting in conformational transition from the CD4-bound state to the unliganded state. In addition, the overlap of the unliganded

gp120 ensemble against the CD4-free gp120 ensemble (18.0%) is smaller than that against the CD4-complex gp120 ensemble (26.9%), revealing that the unliganded gp120 prefer sampling the conformational space sampled by the CD4-complex gp120 to that by the CD4-free gp120. This result seems to be reasonable for reasons that the CD4-independent infection demonstrates the unliganded gp120 is able to fluctuate spontaneously towards the CD4-bound state [17,41], while the CD4-free gp120 inclines to drift away from the CD4-bound conformation [20]. The comprehensively comparative analyses of the essential subspace overlaps between mutated and wild-type gp120 ensembles derived from different starting conformational states indicate that, (i) the unliganded 375 S/W mutant has the largest potential to fluctuate towards the CD4-bound state among all the unliganded models; (ii) the CD4-free 375 S/W mutant possesses the largest capability to maintain the CD4-bound state among all the CD4-free models; (iii) the unliganded 423 I/P mutant has the smallest potential to fluctuate towards the CD-bound state and possesses the largest capability to maintain its unliganded state among all the unliganded models; (iv) the CD4-free 423 I/P mutant possesses the smallest capability to maintain its CD4-bound state. These results allow us to conclude that the 375 S/W mutant samples a conformational space that is very close to the CD4-bound state, while the 423 I/P mutant samples a conformation that is close to the CD4-unliganded state. Our theoretical analytical results are consistent with the experimentally determined mutation effects.

5. Conclusions

Our homology models of the wild-type HIV-1 gp120 in the CD4-complex, CD4-free, and unliganded forms represent three conformational states, namely the excited or pre-fusogenic state with the CD4 docked, the excited state with the CD4 removed, and the relaxed ground state, respectively. The amino acid point mutations were introduced into the wild-type CD4-free and unliganded models to generate the 375 S/W and 423 I/P mutants in the CD4-free and unliganded states, respectively. The CONCOORD computer simulations and subsequent ED analyses for the seven ensembles revealed that the large-scale concerted motions are dominated by intricately combinational rotations of vortices formed between or within the inner domain, outer domain, bridging-sheet and V3 loop. These motional modes will open/close the CD4 cavity and reposition certain structural components that are involved in receptor association or release. Although the large-scale motional modes described by the same ranked eigenvectors in all the simulated models are similar, the variations of the moving directions and motional amplitudes of certain mobile units (or vortices) caused by point mutants would be related to the conformational transition from one state to the other or endow some models with capability to retain their current conformations. The comparative analyses of essential subspace overlaps quantitatively distinguish the preference for conformational transition between gp120 states, and the results are in agreement with chemical intuition: (i) the removal of the CD4 brings gp120 to be unstable and thus the CD4-free gp120 prefers the conformation transition from the CD4-bound state to the unliganded state; (ii) although large structural differences between the unliganded and CD4-free/CD4-complex gp120s, the unliganded gp120 prefers the conformational space of the CD4-complex ensemble to that of the CD4-free gp120 ensemble, providing a plausible explanation for the CD4 independent entry pathway. The comprehensively comparative analyses of the essential subspace overlaps between the mutants and wild-type gp120 ensembles reveal that the 375 S/W mutant, in which the tryptophan indole group occupies the phe43 pocket in the gp120 interior, prefers a conformation close to the CD4-bound state, while the 423 I/P mutant, in which the bridging sheet is hard to be formed, incline to maintain the unliganded conformational state. Our theoretical analyses are in good agreement with the assertion of the gp120 mutation effect based on the experimental observation and MD simulation, therefore the CONCOORD simulation in conjunction with the ED and essential subspace overlap analyses can be extended to a new approach to design or screen mutants that have effects on conformation/function of a protein.

Acknowledgement

The authors thank High Performance Computer Center of Yunnan University for the computational support. This work was supported by Project for Innovative Research Team of High Performance Computing Center in Yunnan University, by grants from the Yunnan University (2004Q013B) and Yunnan Province (2006C008M), by open funds from Laboratory for Conservation and Utilization of Bio-resources and Innovation Group Project from Yunnan Province.

References

[1] F. Barre-Sinoussi, J.C. Chermann, F. Rey, M.T. Nugeyre, S. Chamaret, J. Gruest, C. Dauguet, C. Axler-Blin, F. Vezinet-Brun, C. Rouzioux, W. Rozenbaum, L. Montagnier, Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS), Science 220 (1983) 868–871.

[2] R.C. Gallo, S.Z. Salahuddin, M. Popovic, G.M. Shearer, M. Kaplan, B.F. Haynes, T.J. Palker, R. Redfield, J. Oleske, B. Safai, Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS, Science 224 (1984) 500–503.

[3] M.D. Daniel, N.L. Letvin, N.W. King, M. Kannagi, P.K. Sehgal, R.D. Hunt, P.J. Kanki, M. Essex, R.C. Desrosiers, Isolation of T-cell tropic HTLV-III-like retrovirus from macaques, Science 228 (1985) 1201–1204.

[4] N.L. Letvin, M.D. Daniel, P.K. Sehgal, R.C. Desrosiers, R.D. Hunt, L.M. Waldron, J.J. MacKey, D.K. Schmidt, L.V. Chalifoux, N.W. King, Induction of AIDS-like disease in macaque monkeys with T-cell tropic retrovirus STLV-III, Science 230 (1985) 71–73.

[5] A.G. Dalgleish, P.C. Beverley, P.R. Clapham, D.H. Crawford, M.F. Greaves, R.A. Weiss, The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus, Nature 312 (1984) 763–767.

[6] Y. Feng, C.C. Broder, P.E. Kennedy, E.A. Berger, HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor, Science 272 (1996) 872–877.

[7] A. Trkola, T. Dragic, J. Arthos, J.M. Binley, W.C. Olson, G.P. Allaway, C. Cheng-Mayer, J. Robinson, P.J. Maddon, J.P. Moore, CD4-dependent, antibody-sensitive interactions between HIV-1 and its co-receptor CCR-5, Nature 384 (1996) 184–187.

[8] Q.J. Sattentau, J.P. Moore, F. Vignaux, F. Traincard, P. Poignard, Conformational changes induced in the envelope glycoproteins of the human and simian immunodeficiency viruses by soluble receptor binding, J. Virol. 67 (1993) 7383–7393.
[9] Q.J. Sattentau, J.P. Moore, Conformational changes induced in the human immunodeficiency virus envelope glycoprotein by soluble CD4 binding, J. Exp. Med. 174 (1991) 407–415.
[10] W. Weissenhorn, A. Dessen, S.C. Harrison, J.J. Skehel, D.C. Wiley, Atomic structure of the ectodomain from HIV-1 gp41, Nature 387 (1997) 426–430.
[11] P.D. Kwong, R. Wyatt, J. Robinson, R. Sweet, J. Sodroski, W. Hendrickson, Structure of an HIV-1 gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody, Nature 393 (1998) 649–659.
[12] P.D. Kwong, R. Wyatt, S. Majeed, J. Robinson, R.W. Sweet, J. Sodroski, W.A. Hendrickson, Structures of HIV-1 gp120 envelope glycoproteins from laboratory-adapted and primary isolates, Structure 8 (2000) 1329–1339.
[13] B. Chen, E.M. Vogan, H. Gong, J.J. Skehel, D.C. Wiley, S.C. Harrison, Structure of an unliganded simian immunodeficiency virus gp120 core, Nature 433 (2005) 834–841.
[14] P.D. Kwong, M.L. Doyle, D.J. Casper, C. Cicala, S.A. Leavitt, S. Majeed, T.D. Steenbeke, M. Venturi, I. Chaiken, M. Fung, H. Katinger, P.W. Parren, J. Robinson, D. Van Ryk, L. Wang, D.R. Burton, E. Freire, R. Wyatt, J. Sodroski, W.A. Hendrickson, J. Arthos, HIV-1 evades antibody-mediated neutralization through conformational masking of receptor-binding sites, Nature 420 (2002) 678–682.
[15] D.G. Myszka, R.W. Sweet, P. Hensley, M. Brigham-Burke, P.D. Kwong, W.A. Hendrickson, R. Wyatt, J. Sodroski, M.L. Doyle, Energetics of the HIV gp120-CD4 binding reaction, Proc. Natl. Acad. Sci. U.S.A. 97 (2000) 9026–9031.
[16] R. Wyatt, P.D. Kwong, E. Desjardins, R.W. Sweet, J. Robinson, W.A. Hendrickson, J. Sodroski, The antigenic structure of the HIV gp120 envelope glycoprotein, Nature 393 (1998) 705–711.
[17] S.H. Xiang, P.D. Kwong, R. Gupta, C.D. Rizzuto, D.J. Casper, R. Wyatt, L.P. Wang, W.A. Hendrickson, M.L. Doyle, J. Sodroski, Mutagenic stabilization and/or disruption of a CD4-bound state reveals distinct conformations of the human immunodeficiency virus type 1 envelope glycoprotein, J. Virol. 76 (2002) 9888–9899.
[18] H.J.C. Berendsen, S. Hayward, Collective protein dynamics in relation to function, Curr. Opin. Struct. Biol. 10 (2000) 165–169.
[19] S.T. Hsu, A.M.J.J. Bonvin, Atomic insight into the CD4 binding-induced conformational changes in HIV-1 gp120, Proteins 55 (2004) 582–593.
[20] Y. Pan, B. Ma, O. Keskin, R. Nussinov, Characterization of the conformational state and flexibility of HIV-1 glycoprotein gp120 core domain, J. Biol. Chem. 279 (2004) 30523–30530.
[21] Y. Pan, B. Ma, R. Nussinov, CD4 binding partially locks the bridging sheet in gp120 but leaves the beta2/3 strands flexible, J. Mol. Biol. 350 (2005) 514–527.
[22] B.L. de Groot, D.M.F. van Aalten, R.M. Scheek, A. Amadei, G. Vriend, H.J.C. Berendsen, Prediction of protein conformational freedom from distance constraints, Proteins 29 (1997) 240–251.
[23] C.P. Barrett, M.E. Noble, Molecular motions of human cyclin-dependent kinase 2, J. Biol. Chem. 280 (2005) 13993–14005.
[24] C.P. Barrett, B.A. Hall, M.E. Noble, Dynamite: a simple way to gain insight into protein motions, Acta Crystallogr. D Biol. Crystallogr. 60 (2004) 2280–2287.
[25] L.V. Mello, B.L. de Groot, S. Li, M.J. Jedrzejas, Structure and flexibility of Streptococcus agalactiae hyaluronate lyase complex with its substrate. Insights into the mechanism of processive degradation of hyaluronan, J. Biol. Chem. 277 (2002) 36678–36688.
[26] J. Vreede, M.A. van der Horst, K.J. Hellingwerf, W. Crielaard, D.M. van Aalten, PAS domains. Common structure and common flexibility, J. Biol. Chem. 278 (2003) 18434–18439.
[27] A. Bairoch, R. Apweiler, C.H. Wu, W.C. Barker, B. Boeckmann, S. Ferro, E. Gasteiger, H. Huang, R. Lopez, M. Magrane, M.J. Martin, D.A. Natale, C. O'Donovan, N. Redaschi, L.S. Yeh, The Universal Protein Resource (UniProt), Nucl. Acids Res. 33 (2005) 154–159.
[28] N. Deshpande, K.J. Addess, W.F. Bluhm, J.C. Merino-Ott, W. Townsend-Merino, Q. Zhang, Describes the capabilities of the PDB Beta site, Nucl. Acids Res. 33 (2005) 233–237.
[29] W.F. Vranken, M. Budesinsky, F. Fant, K. Boulez, F.A. Borremans, The complete consensus V3 loop peptide of the envelope protein Gp120 of HIV-1 shows pronounced helical character in solution, FEBS Lett. 374 (1995) 117–121.
[30] A. Sali, T.L. Blundell, Comparative protein modelling by satisfaction of spatial restraints, J. Mol. Biol. 234 (1993) 779–815.
[31] A. Amadei, A.B.M. Linssen, H.J.C. Berendsen, Essential dynamics of proteins, Proteins 17 (1993) 412–425.
[32] H.J.C. Berendsen, D. Van Der Spoel, R. Van Drunen, GROMACS: a message-passing parallel molecular dynamics implementation, Comp. Phys. Commun. 91 (1995) 43–56.
[33] W. Humphrey, A. Dalke, K. Schulten, VMD—visual molecular dynamics, J. Mol. Graph. 14 (33–38) (1996) 27–28.
[34] A. Merlino, L. Vitagliano, M.A. Ceruso, L. Mazzarella, Dynamic properties of the N-terminal swapped dimer of ribonuclease A, Biophys. J. 86 (2004) 2383–2391.
[35] B.L. de Groot, S. Hayward, D.M.F. van Aalten, A. Amadei, H.J.C. Berendsen, Domain motions in bacteriophage T4 lysozyme: a comparison between molecular dynamics and crystallographic data, Proteins 31 (1998) 116–127.
[36] J.D. Faraldo-Gómez, L.R. Forrest, M. Baaden, P.J. Bond, C. Domene, G. Patargias, J. Cuthbertson, M.S.P. Sansom, Conformational sampling and dynamics of membrane proteins from 10-nanosecond computer simulations, Proteins 57 (2004) 783–791.
[37] B. Hess, Similarities between principal components of protein dynamics and random diffusion, Phys. Rev. E 62 (2000) 8438–8448.
[38] C.C. Huang, M. Tang, M.Y. Zhang, S. Majeed, E. Montabana, R.L. Stanfield, D.S. Dimitrov, B. Korber, J. Sodroski, I.A. Wilson, R. Wyatt, P.D. Kwong, Structure of a V3-containing HIV-1 gp120 core, Science 310 (2005) 1025–1028.
[39] R.A. Laskowski, M. MacArthur, D.S. Moss, J.M. Thornton, PROCHECK: a program to check the stereochemical quality of protein structures, J. Appl. Crystallogr. 26 (1993) 283–291.
[40] D. Baker, A. Sali, Protein structure prediction and structural genomics, Science 294 (2001) 93–96.
[41] P. Kolchinsky, T. Mirzabekov, M. Farzan, E. Kiprilov, M. Cayabyab, L.J. Mooney, H. Choe, J. Sodroski, Adaptation of a CCR5-using, primary human immunodeficiency virus type 1 isolate for CD4-independent replication, J. Virol. 73 (1999) 8120–8126.