Fluctuation dynamics analysis of gp120 envelope protein reveals a topologically based communication network

Indira Shrivastava¹ and Judith M. LaLonde²

¹ Department of Computational Biology, School of Medicine, University of Pittsburgh, 3083 Biomedical Science Tower 3, 3501 Fifth Avenue, Pittsburgh, PA 15213
² Chemistry Department, Bryn Mawr College, 101 N. Merion Ave., Bryn Mawr, PA 19010

ABSTRACT

Human Immunodeficiency Virus (HIV) infection is initiated by binding of the viral glycoprotein gp120, to the cellular receptor CD4. On CD4 binding, gp120 undergoes conformational change, permitting binding to the chemokine receptor. Crystal structures of gp120 ternary complex reveal the CD4 bound conformation of gp120. We report here the application of the Gaussian network model (GNM) to the crystal structures of gp120 bound to CD4 or CD4 mimic and 17b, to study the collective motions of the gp120 core and determine the communication propensities of the residue network. The GNM fluctuation profiles identify residues in the inner domain and outer domain that may facilitate conformational change or stability, respectively. Communication propensities delineate a residue network that is topologically suited for signal propagation from the Phe43 cavity throughout the gp120 outer domain. These results provide a new context for interpreting gp120 core envelope structure–function relationships.

INTRODUCTION

The Human immunodeficiency virus (HIV) causes the development of acquired immunodeficiency syndrome (AIDS) by depletion of CD4+ lymphocytes of an infected individual.¹² The infection is mediated by a series of attachment events initiated by the HIV viral coat glycoprotein, gp160 which is cleaved into its two components, gp120 and gp41.³ The HIV glycoproteins, gp120 and gp41, are assembled as a trimer.⁴,⁵ The infection process of HIV in human T-cell lymphocytes occurs via binding of gp120, to the host T-cell CD4 receptor⁶,⁷ followed by gp120 restructuring.⁸,⁹ This conformational change exposes the binding site of the chemokine receptor, allowing binding of gp120 to either CCR5 or CXCR4. Chemokine binding is the second obligatory event to viral entry¹⁰–¹³ and is followed by insertion of the gp41 fusion-peptide in the host cell membrane allowing fusion and viral entry.¹⁴–¹⁷

Several X-ray structures of the gp120:CD4:17b-antibody complex elucidate the conformation of gp120 after binding CD4 and 17b, ¹⁸,¹⁹ (Fig. 1). These X-ray structures of the CD4 bound gp120 form, reveal three key domains—the inner domain, outer domain and a bridging sheet—that fold to form a large binding cavity. Two key CD4 residues, Phe43 and Arg59, bind in the gp120 cavity ¹⁸ and to Asp 368 on an adjacent alpha-helix, respectively. However, when Phe43 and Arg59 are mutated to Ala, CD4 no longer binds gp120.²⁰ The core gp120 protein from various strains has also been crystallized with several antibodies,²¹ revealing the structure of the previously undescribed third variable (V3) loop. In addition, several mini-protein CD4 mimetics have also been solved, bound to gp120²² showing a biphenyl group bound deep in the CD4 cavity. A recent structure of gp120 containing the gp41 interacting region²³ when compared with the CD4-binding site antibody, F105–gp120 complex²⁴ with a disassociated bridging sheet, reveals the plasticity of the inner and bridging sheet domains. Furthermore, modification of CD4 Phe43 with various derivatized bromoacetamide conjugates indicates that the gp120 cavity has the flexibility to accommodate ligands much larger than the CD4 Phe43.²⁵ Although a crystal structure of the...
unliganded HIV gp120 does not exist, that of the unbound Simian immunodeficiency virus (SIV) gp120 has been solved. The structure of SIV gp120, which has 35% sequence identity with HIV gp120, indicates an invariant outer domain, with conformational changes occurring in both the bridging sheet and inner domain.

Thermodynamic studies indicate that CD4 binding to gp120 results in a highly favorable binding enthalpy ($\Delta H = -63 \text{ kcal/mol}$) balanced with a highly unfavorable molecular ordering ($-\Delta TS = 52 \text{ kcal/mol}$). The large conformational change may be accounted for by the restructuring of up to 100 amino-acid residues or the burying of 10,000 $\text{Å}^2$ of surface area. Furthermore, titration calorimetry from two mutant gp120 envelope proteins (S375W and I432P) suggests two distinct conformational states, a CD4-bound state represented by the crystal structure of the CD4-gp120-17b complex and a S375W mutant and a non-CD4 bound state represented by an I423P mutation. A third cavity stabilizing mutation T257S has been characterized and the double mutant T257S/S375W has increased CD4 affinity. The conformational change is thought to play a role in formation of the cytokine receptor binding interface, which spans portions of both gp120 and CD4. A series of compounds (denoted NBD) discovered by Zhao et al. in database screening have been shown to induce the CD4 stimulated conformational change in a manner similar to CD4 binding and enhance viral infection on CD4-deficient target cells. These compounds compete with CD4 binding to gp120 and enhance binding of CD4-gp120 to the chemokine receptor CCR5. Mutational analysis of gp120 cavity residues has shown that several mutations increased NBD compound affinity and enhanced viral infectivity on CD4 deficient cells. Furthermore, second generation compounds with improved affinity have shown to cause a rapid inactivation of the virus, establishing it as a potential entry inhibitor and antiviral therapeutic agent.

The dynamics of CD4 binding to gp120 has a higher mobility, aiding the CD4 molecule in binding to target molecules. The dynamics of gp120 proteins also have been examined computationally. Molecular dynamics (MD) simulations of wild-type gp120 and S375W mutant indicated significant differences in protein flexibilities. The S375W mutant form preferred the bound-like conformation while the wild-type diverged from this conformation, with partial unfolding of some of the $\beta$-strands. Nonequilibrium steered MD, in which the bridging sheet $\beta$-2/3 or $\beta$-20/21 were pulled away from an $\alpha$-helix-1 of the inner-domain indicated that the $\beta$-2/3 had more flexibility than the $\beta$-20/21, which favored interaction with the inner-domain $\alpha$-helix, keeping a conformation close to the CD4-bound-form. Liu et al. examined the dynamic domains of homology models of HIV-1 gp120 core, in the presence and absence of CD4 and the effects of mutation on these motions. They generated an ensemble of conformations for different gp120 models and performed an essential dynamics analysis to identify the principal modes of motion. Interestingly, their results also indicated that the S375W mutation favors the CD4 bound conformation, whereas the I423P conformation prefers the unliganded conformation. MD studies by Hsu et al. indicate that there are concerted loop motions in the vestibule of the CD4 cavity, stabilization of the bridging sheet and a coalescing of the bridging sheet and V3 loop to form the coreceptor binding site. Binding entropies extracted from these MD trajectories suggest that the large entropy loss associated with CD4 binding is derived from hydrophobic interactions from CD4 Phe43 insertion into the cavity, the formation of a hydrogen bonding network and the restructuring of the bridging sheet. Temperature accelerated MD was used for large-scale sampling of gp120 motions by Abrams and Van den-Eijnden and predicted a counter rotation between the inner and outer domains as well as disruption of the bridging sheet in the unbound form of gp120. Tan and Rader have used a program that implements graph theory to analyze the flexibility and rigidity of all known gp120 structures. A flexibility index describing the extent and distribution of flexible and rigid regions in the inner, outer, and bridging sheet domains of the 22 gp120 structures with various ligands indicated the inner domain and bridging sheet domains to be more flexible, whereas the outer do-
main to be more rigid. Gp120 proteins bound with CD4 exhibited less flexibility in the inner domain than gp120 bound to mini-protein mimetics. Increasing inner domain flexibility accompanied by increasing outer domain rigidity was also observed for the b12 antibody bound gp120 complex. Comparison of these various structures (CD4 bound, b12 bound, and apo gp120) led to identification of a universal rigid region on a $\alpha$-2 helix (residues 335–352) which is proposed to be a potential initial recognition site associated with coreceptor binding. A consensus rigid cluster on a $\beta$-sheet located on the coreceptor binding surface was also identified. 41

Another coarse-grained approach to protein flexibility is the Gaussian network model (GNM) 44–46 which permits an efficient exploration of collective motions of proteins. The GNM approach is based on local packing density and bonded and nonbonded contact topology in a given structure. The decomposition of the vibrational modes elucidates the slowest (global motion) and the fastest (local motion) modes. The slowest mode provides information on the large-scale cooperative movements of large domains, and the regions with minimal motion during these collective motions. The regions with restricted motion are considered to play an important role in modulating and monitoring the catalytic activity of enzymes and serve as focal points for modulating collective domain motions or hinge bending.44,47,48 GNM has been shown to yield results which are in remarkable agreement with results obtained from models with standard semiempirical potentials, as far as low-frequency normal modes are concerned.49–51 The collective motions also determine the communication propensities inherent to the protein structure. The communication propensities of residues in a given network can be expressed in terms of the Kirchoff matrix of inter-residue contacts, which is also the underlying theory of GNM.51 In this network model, the catalytic residues are distinguished by their fast and precise communication capabilities.51

In this study, we present the dynamics and communication propensities of the gp120 core residues calculated with GNM, to provide a computational framework for elucidating the residues that are critical in stabilizing the unbound form of gp120 and that are key communicators of gp120 conformational change. Here, we report the results for three CD4 bound-gp120 crystal structures: the gp120, CD4, 17b complex from HIV strain HXBC2 (PDB ID: 1G9M)19; the gp120, CD4, 17b complex from HIV strain YU2 (PDB ID: 1G9N),19 and the gp120, 43 residue scorpion-toxin CD4 mimic, and 17b complex from HIV strain YU2 (PDB ID: 2I5Y).22 By analysis of gp120 fluctuation dynamics and communication propensities, a stable outer domain is revealed with a communication network emanating from the Phe43 cavity and poised for signal propagation throughout the gp120 outer domain.

RESULTS AND DISCUSSION

GNM fluctuation profiles

We have adopted a simplified nomenclature to represent the various complexes and components used in this study (Table 1). Briefly, for the HXBC2 strain, the three component complex of gp120, CD4, and 17b is referred to as COM1, the two component complex of gp120 and CD4 is referred to as GCD1, and gp120 alone is referred to as GPO1. Similarly, complexes from the YU2 strain are denoted COM2, GCD2, GPO2 and the third system with the scorpion-toxin CD4 mimic bound to Gp120 and 17b are labeled COM3, GCD3, and GPO3. Residue numbering for the core gp120 crystal structures maintains the numbering from the full-length gp120 sequence (see Materials and Methods section). We first sought to establish agreement between experimental Debye-Waller temperature factors (or B-factors) and the theoretical B-factors, computed from GNM, for the three systems of study. B-factors provide an experimental measurement of the intrinsic protein dynamics52 which is defined by the local packing density and is often used to establish a correlation between theory and experiment.53,54 As shown in Figure 2, there is a reasonable agreement between the experimental and theoreti-
cal B-factors for gp120 in COM3, GCD3, and GPO3. The broad flat peaks correspond to the deletions of the V1/V2 (residues 129–194) and V3 (residues 300–329) loops in the gp120 core X-ray structure. This overall agreement between X-ray and GNM B-factors is seen in the other gp120 complexes as well (see Supporting Information, Figures S1A, and S1B) and is consistent with earlier observations on other protein systems.45–48,55

The second question we posed was does the gp120 core exhibit a different fluctuation profile for the two strains, HXBC2 and YU2? The YU2 and HXBC2 sequences of the core region of gp120s have 86% sequence identity.19 Most of the sequence differences observed are in the outer domain and are listed in Table S1 of the Supporting Information. A total of 33 outer domain residues in the region from V275 to E464 show conservative amino acid sequence substitutions.

The slow-mode profiles in GNM provide information on the intrinsic dynamics of the protein, with the maxima considered as recognition sites and the minima as hinge sites.48 We focus our analysis on the slowest GNM modes which will provide information on the cooperative motions of gp120 domains, to elucidate the components of gp120 flexibility. Figure 3A depicts the slow-mode profile (or mode shape) for GPO1, GPO2, and GPO3, as a function of the residues. The ordinate represents the normalized mean square (ms) fluctuation of residues, driven by the slowest two modes (2<, k < 3) of motion. Although the maxima in GPO3 are less pronounced than in GPO1 or GPO2, the overall profiles are similar, with both minima and maxima occurring in the same vicinity, for all the three structures. Table S2 presents the residues corresponding to minima and maxima in GPO3. Despite the modest differences in the amino-acid sequence between HXBC2 and YU2 gp120 proteins, the cores exhibit similar dynamics (Fig. 3A). This result is consistent with the similar binding affinities and energetic profiles determined by isothermal titration calorimetry and surface plasmon resonance respectively.27,28,31,56

We considered two complexes from the YU2 strain with different primary ligands, CD4 and a 43 residue mini-protein derived from scorpion-toxin (GPO2 and GPO3, respectively). The gp120 crystal structures from these complexes differ by a C-alpha RMSD of 1.55 Å. The normalized mean square fluctuation plots of the two gp120 cores show the same fluctuation profile, despite the difference in primary ligand in the crystal structure complex. The similarity of GNM slow-mode profiles regardless of HIV strain and primary ligand indicates a set of conserved collective motions common to all monomeric core gp120 envelope proteins. The residues, at which the GNM slow-mode profiles exhibit minima in

Table II
Residues in GPO3 (gp120 from 2I5Y) that Have GNM Calculated Minima

<table>
<thead>
<tr>
<th>Residue</th>
<th>Domain</th>
<th>2nd Structure</th>
<th>Residue</th>
<th>Domain</th>
<th>2nd Structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>T232</td>
<td>Inner</td>
<td>Loop A</td>
<td>P470</td>
<td>Outer</td>
<td>β-24</td>
</tr>
<tr>
<td>N234</td>
<td>Inner</td>
<td>Loop A</td>
<td>G471</td>
<td>Outer</td>
<td>β-24</td>
</tr>
<tr>
<td>R252 to V255</td>
<td>Inner</td>
<td>Loop B</td>
<td>G472</td>
<td>Outer</td>
<td>β-24</td>
</tr>
<tr>
<td>S256 to T257</td>
<td>Outer</td>
<td>Loop B</td>
<td>G473</td>
<td>Outer</td>
<td>β-24</td>
</tr>
<tr>
<td>O258</td>
<td>Outer</td>
<td>Loop B</td>
<td>D474</td>
<td>Outer</td>
<td>β-24</td>
</tr>
<tr>
<td>L259 to R273</td>
<td>Outer</td>
<td>Loop B, B-9, Loop 10</td>
<td>N475</td>
<td>Inner</td>
<td>α-5</td>
</tr>
<tr>
<td>I285 to S291</td>
<td>Outer</td>
<td>B-11</td>
<td>D477</td>
<td>α-5</td>
<td></td>
</tr>
<tr>
<td>E370</td>
<td>Outer</td>
<td>α-3</td>
<td>N478</td>
<td>Inner</td>
<td>α-5</td>
</tr>
<tr>
<td>I371</td>
<td>Outer</td>
<td>α-3</td>
<td>S481 to E482</td>
<td>Inner</td>
<td>α-5</td>
</tr>
<tr>
<td>S375 to N377</td>
<td>Outer</td>
<td>B-16</td>
<td>F383</td>
<td>B-17</td>
<td></td>
</tr>
<tr>
<td>F383</td>
<td>Outer</td>
<td>B-17</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I499 to L453</td>
<td>Outer</td>
<td>B-22, B-23</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Residues depicted in bold, italic and underlined fonts are conserved among all primates strains, all human strains and moderately conserved in human strains, respectively.

*secondary structure.
the three gp120s, are tabulated in the Table 2. We choose to focus the description of the conserved collective motions using the coordinate set for GPO3, because it represents a higher resolution structure from the YU2 strain and the biphenyl group binds more deeply in the gp120 cavity. GPO3 minima are depicted in Figure 4 and
listed in Table 2. As illustrated in Figure 4A, most of the minima occur at the interface between the inner and outer domain. Many of the residues that display minima in the GNM mean square fluctuations are also within 4.5 Å of W427 in the Phe43 cavity (Fig. 3). These residues span a portion of the Phe43 cavity, the interface between the inner domain (α-helix-5 and loop A) and outer domain (β-strands 9, 10, 11, 23, 24, and loop outer domain α-helix 3. The majority of observed minima are in the outer domain (Fig. 4A) with fewer minima in the inner domain (Fig. 4A). Notably, none of the residues from the bridging sheet domain display minima in GNM slow modes. The residues at minima occupy critical loci in the global mode and could be viewed as trigger points for controlling gp120 conformational change. These sites may function either in a hinge binding capacity or define a core folding nucleus found in the prebound state of unliganded HIV gp120. We note that the core folding nucleus described here, does not imply a folding mechanism, but rather represents a set of residues that may be loosely structured in various prebound states, suggesting that other folding events may crystallize around this nucleus. The crystal structure of the unliganded SIV gp12026 indicates that much of the outer domain shares a common structure with the HIV CD4 bound conformation. Thus, it is reasonable to conclude that minima occurring in the outer domain represent a core folding nucleus found in the prebound form of gp120. Furthermore, this conclusion is supported by the observation that most of the residues which occur at a minimum in GNM slow mode, are also conserved across primate and human HIVs (Q258, E370, I371, P470, G471, G472, G473, D474, M475 and D477) (Table 2). Of the eight gp120 residues that are within 4.5 Å of the miniprotein biphenyl in the 2I5Y complex, (T257, E370, I371, S375, N425, M475, W427, and G473) six display minima, T257, E370, I371, S375, M475, and G473 (all from the outer domain). Interestingly, a residue that does not display a GNM slow-mode minimum is W427. This residue from the bridging sheet is highly conserved across all primate HIV sequences, lines the ligand-binding pocket (Phe43 cavity) and forms hydrophobic interactions with CD4-Phe43 in the gp120-CD4 complexes and the biphenyl in CD4M47 receptor.18,19,21 Residues R252 to E275, in the low-fluctuation region, span the inner domain and the outer domain (Fig. 4B). Within this span resides the tripeptide, L259–L260–L261 adjacent to the Q258 which is conserved in HIV. A second region of low fluctuation occurs from I449 to L453 which includes a portion of the tripeptide L452–L453–L454. A third low-fluctuation region, from residues P470 to M475, contains the tripeptide G471–G472–G473. These three regions are noteworthy because they contain the conserved residues Q258, P470, G471, G472, G473, and M475. This core region of GNM, including regions of low fluctuation and slow-mode minima, extends to β-strands on both sides encompassing residues I285–S291, S375–N377 and F383. As viewed in Figure 4C, the GNM minima form a continuous surface of a stable secondary structure. Interestingly, the GNM minima are only located on one face of the Phe43 cavity, suggesting that this face provides a core structural region that is similarly structured in the CD4 unbound form of gp120.

Several GNM slow-mode minima are also exhibited by residues in the inner domain, H105, E211, P212, I213, K232, N234, R252–V255, M475, D477, N478, N481, and E482. Of these, H105 and M475 are conserved residues and are located in the inner domain on α-helix 1 and 5, respectively (Fig. 4D). Both these residues interact with the highly conserved W427 in the Phe43 cavity. In the X-ray structure, W427 is closely interacting with a conserved residue, I109 which is one-turn away from H105 on α-helix 1. These residues (H105 and M4745) seem to stabilize the Phe43 cavity by maintaining tight interactions with W427. In addition, M475 also interacts with G473 of the G471–G471–G473 tripeptide, in the outer domain. Thus, it is tempting to speculate that H105 and M475 serve as hinge residues that help drive the conformational change that forms the Phe43 cavity. Notably, many of the residues initially characterized by mutagenesis as essential for gp120-CD4 binding5,57,58 are observed to either be at a minimum or in its close proximity in GNM slow mode, in this study (S256–T257, L259, A266–E267, E269, E370, Y384, P470, D477, and E482).

Although the residues at minima are known to be critical in catalytic or binding sites, those at the maxima have been observed to be good recognition sites.55 The entire loop between bridging sheet strands β-20 and β-21(residues N425 to K432) exhibits a maximum in the GNM slow-mode profile, underlying the significance of this region as a recognition site for CD4 binding. This includes the highly conserved W427 which forms key hydrophobic interactions with the biphenyl of mini-protein (GCD3) (Fig. 4D) and CD4 Phe43 (GCD1 and GCD2). The position of W427 on the bridging sheet in the Phe43 cavity suggests that it might indeed be playing a critical role in recognition of the CD4 receptor, and has in fact been known to be critical in receptor–protein interactions.57 The less-conserved N425 (a maximum) on β-sheet 20 of the bridging sheet interacts with the highly conserved E370 (a minimum) (Fig. 4E), a GNM minima adjacent to the Phe43 cavity in the outer-domain. These two residues together provide rigidity (minimum) as well as flexibility (maximum), at the interface of the bridging sheet and outer domain. This juxta-positioning of a rigid region with a flexible region is ideal for ligand binding; the flexible region enhances ligand-recognition, whereas the rigid-region enhances ligand-binding stability. The residues N425 and W427 are in direct contact with the CD4-Phe43 ligand, in the complexed structure (GCD2), and with the mini-protein biphenyl (GCD3), thus further supporting their importance as recognition sites.
A preliminary energy minimization, followed by a short MD simulation in a solvated environment, is expected to relax the X-ray structure, driving it closer to its true minimum. Thus, we performed MD simulations of solvated GPO3 structure and analyzed the dynamics. (See Supporting Information for simulation protocols). Five nanosecond (ns) simulation trajectories for GPO3 were generated, using GROMACS59,60 in a fully solvated environment, at 310K and under constant NPT conditions. The root mean square fluctuations (rmsf) of the GP120 from MD simulations are in excellent agreement with the mean square fluctuations observed with GNM (Supporting Information Fig. S2). The former is the mean square fluctuation from the 5 ns average of the simulation trajectory, whereas the latter is the mean square fluctuation (B-factor), calculated with GNM, for the mean structure, averaged over 5-ns simulation trajectory. Interestingly, these two quantities also agree well with the experimentally determined B-factors of the GPO3 X-ray structure (Supporting Information Fig. S2).

As shown in Figure 3B, the GNM mean square fluctuation profile for the MD relaxed GPO3 coordinate set closely mirrors the profile from the GPO3 X-ray structure. Comparing the GNM fluctuation profiles produced from MD snapshots for GPO3 and GCD3 and GCD2, Supporting Information S3, also indicates that MD relaxation of the gp120-CD4 (or miniprotein) bound complexes reproduces a fluctuation profile similar to the MD relaxed GPO3 (relaxed in the absence of ligand). The GNM dynamics calculated with a simple, Hookean potential are thus indeed robust and comparable with that calculated from an atomic level force-field potential, used in MD simulations, suggesting that relaxation of an X-ray structure by MD simulations does not significantly alter the GNM dynamics.

Communication propensities

The GNM slow-mode profiles have elucidated regions of fluctuation minima and maxima in the gp120 proteins. As noted by Chennubhotla and Bahar,51 equilibrium motions exhibited by the protein also determine the communication mechanism inherent in the protein residue network. Moreover, residues which display a high coordination number in a given protein network, are also effective communicators, providing a topological basis for communication propensities (CP) of residues in a protein. This communication propensity analysis, based on elastic network models, directly relates residue fluctuation to their CPs,51,61 i.e., residues whose distances fluctuate with low intensity communicate with a higher efficiency than residues with larger fluctuations. In Figure 5, the commute times for GPO1, GPO2, and GPO3 are plotted. We choose to use a cutoff value of commute times \( C_{ij} < 0.21 \) as criteria for a residue pair \((i,j)\) which are at a distance of at least 10 Å or more, as having good CPs. Although there is no direct correlation between commute times and physical distances, there are some residue pairs which communicate efficiently, despite their long physical distances. As previously noted, the GNM slow modes of GPO1, GPO2, and GPO3 display a similar dynamic profile. The communication profiles of GPO1, GPO2, and GPO3 (Fig. 5) also exhibit a similar pattern of residue communication propensities. Inspection of Figure 5 reveals differences in communicating pairs of residues that are concentrated in the inner domain and bridging sheet domains (the region spanning residues 80–230 in Figure 5) for GPO1, GPO2, and GPO3. The amino-acid sequence as well as secondary and tertiary structures are conserved between HXBC2 (GPO1) and YU2 (GPO2 and GPO3) strains. Tan and Rader41 also observed a variation in the domain flexibility among HIV strain and bound CD4 ligands in a study of relative domain flexibility for twenty CD4 and scyllatoxin CD4 mimic-gp120 crystal structures. In a broad sense, the set of conserved communication hubs identified in this study serve as a basis for hypothesis generation and further experimentation. Furthermore, recent structure reports23,62 indicate that the inner domain contains three structurally distinct layers that define gp120 mobility accounting for the scarcity of GNM slow-mode minima and residues with good communication propensities in this region.

Of the 234 communicating pairs with \( C_{ij} \) values <0.21 in the three structures studied (Supporting Information Table 3) 47% are conserved between GPO1, GPO2, and GPO3 and another 20% are conserved among two of the three structures. In this analysis, we focus on
Fluctuation Profiles are Sorted by Communication Hub in the Bridging Sheet, Inner Domain and Outer Domain

<table>
<thead>
<tr>
<th>Hub (i)</th>
<th>Hub (j) 1</th>
<th>Hub (j) 2</th>
<th>Hub (j) 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bridging Sheet</td>
<td>Inner Domain</td>
<td>Outer Domain</td>
<td></td>
</tr>
<tr>
<td>A219 I224 I225 K227 L226* C228*</td>
<td>A224 I225</td>
<td>T244 V242*</td>
<td>K487* V488</td>
</tr>
<tr>
<td>C239 V242</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L265* A266*</td>
<td>L270* V271* I272</td>
<td>A270</td>
<td></td>
</tr>
<tr>
<td>I270*</td>
<td>I272</td>
<td>C239</td>
<td></td>
</tr>
<tr>
<td>I278*</td>
<td>I285</td>
<td>I294*</td>
<td>N295 N332</td>
</tr>
<tr>
<td>I284*</td>
<td>V286*</td>
<td></td>
<td>S447* N448</td>
</tr>
<tr>
<td>I294*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N295</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C296</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H330</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H331* C331* N332 L333</td>
<td>I414 L416* L418</td>
<td>I466 I467 F468* R469*</td>
<td></td>
</tr>
<tr>
<td>I359 I360* F361 N362*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F361 N362 N363</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H371 T373</td>
<td>H374* S375* F376</td>
<td>F383 Y384* C385* N386</td>
<td></td>
</tr>
<tr>
<td>H374* S375* F376</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F383 Y384* C385</td>
<td>L416 C418 I420</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L453 L454 T455* R456</td>
<td>F468 R469* P470</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L453 L454 T455* R456</td>
<td>F468 R469* P470</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L453 L454 T455* R456</td>
<td>F468 R469* P470</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Residues that also exhibit GNM slow mode minima or maxima are in bold or underlined font, respectively. Residues that are conserved in HIV sequences are indicated in italics (G258, P270, R469, G471, G472, T458. *Residues pairs with efficient communication propensities as determined by MD simulation. The amino-acid residue is given for YU2/GP03 only.

The conserved set residue of pairs with low communication propensity observed in GPO1, GPO2, and GPO3 and listed in Table 3. Residues pairs with low communication propensity for GCD1, GCD2, GCD3, COM1, COM2, and COM3 are tabulated in Supporting Information Tables S4 and S5, respectively, and did not indicate a significant difference in intradomain gp120 communication. Pairs of good communicators between the inner, outer, and bridging domains are not observed in any of the three gp120 structures, GPO1, GPO2, and GPO3. Communication hubs in gp120 are only observed between intradomain residues. In the bridging sheet domain, one communication hub exists between residues P118, C118, K121, L122 on β-strand 2 and adjacent β-strand 22. In the inner domain, residues A224–C228 form good communication with residues V242, T244, V245, and also K487 and V488. This set of residues resides in the β-sheet portion of the inner domain. However, the most extensive network of communicating hubs occurs in the outer domain. Many of these outer domain residue pairs with low commute times and which also display GNM slow-mode minima are residues within β-strand secondary elements (Table 3). However, only two residues with low commute times (V242 and T244) exhibit GNM slow-mode maxima. The dominant communication hub in the outer domain is anchored by residues T257–N262 with low commute times to one hub at residues H374–F376 and a second hub at residues I449–L454, and a third at residues P470–G472. Other core hubs of communication occur with residues Q287–N289 and I270–I272. The full web of interconnecting residues pairs for outer domain residues can be delineated from residues in Table 3 and is shown via cartoon diagram in Figure 6.

To corroborate this set of communication hubs, we examined all atom MD simulation of GPO3. The minimum distances between C–α atoms of communicating hubs was calculated from a 17 ns trajectory (Supporting Information Fig. 4). These minimum distances are maintained over the course of the MD simulation indicating that these residues indeed exhibit low mobility, and that signal propagation and equilibrium fluctuations are related. A comparison of the cross-correlation matrices for GNM fluctuations, GNM communication propensities, and the motions of C–α atoms around the averaged position from MD simulation (Fig. 7) shows similarities between regions of low communication propensity with both GNM and MD correlated motions. Several regions with low communication propensities suggested as communication hubs display a positive correlation of C–α motion, suggesting that the local packing density is reflected in the communication propensities, albeit, indirectly. We note that there is no one to one correspondence between slow-mode minima and communication propensity. Although most of the residues that are minima in a given slow mode, may communicate efficiently, there are also residues which have a high mobility, and yet are good communicators, because of the positive correlation between the fluctuations. The commute times reflect the correlation between the fluctuations in the inter-residue distances, whereas the minima
in a given slow mode reflect low mobility at a given residue, in that particular slow mode.

To further elaborate the communication network based on MD simulation, we calculated residue communication propensity from the trajectory of GPO3 as defined by Morra et al. Residues with low communication propensities are listed in Supporting Information Table 3. The two methods reveal a set of overlapping residues (Table 3) that communicate efficiently. Because the two methods sample protein motion of different magnitudes (0.1 nm vs. 1 nm), agreement between the two sets reveals the most efficient residue communicators (Table 3). As shown in Figure 6, GNM determined CP define broader regions of communicating pairs of residues, whereas the MD communication propensities provide detail on which residues pairs communicate the most efficiently.

Two residues identified by both methods as good communicators which are located within 4.5 Å of the biphenyl group of the miniprotein are T257 and S375. Furthermore, we note a correspondence between residues that have been identified as good communicators and have been shown by mutagenesis to be essential for CD4 binding are residues V120, K121, K227, T257, L259, N262 and Y384, (these residues all participate in key communication hubs). Mutation of the residues to alanine may affect the local packing density which may in turn influence the propensity to communicate efficiently with neighboring residues. Analysis of the cross-correlation matrix of the C-α atomic motion from the average structure produced from the GPO3 MD trajectory (Fig. 7C) confirms a positive correlation in motion between several of the Phe43 cavity residues (371 to 383, 386 to 330-330 to 414-418) supporting the notion of signal propagation from the Phe43 cavity to adjacent regions of the outer domain.

**DISCUSSION**

We have used GNM, MD, and the Markovian stochastic model of information diffusion to analyze the equilibrium fluctuations and the communication network of the HIV viral envelope protein, gp120, using the existing X-ray structures available in the CD4 bound form. The gp120 core demonstrates a large conformational change that accompanies molecular recognition of the CD4 and chemokine receptors. The rationale for focusing computation on the CD4 bound form is based on the observation that the NBD class of small molecules drives the structuring of gp120 to the same extent as CD4 receptor binding.

We aimed to analyze the equilibrium fluctuations and signal propagation as a means of understanding which amino acid residues play a role in signal transduction in gp120 with the intent of delineating important components of the prestructured regions of gp120 in the CD4 bound form. GNM slow-mode profiles identified key residues with low mean square fluctuations, in the inner and outer domains. The residues at minima in the outer domain are hypothesized to form a core folding nucleus, whereas those in the inner domain serve as hinge sites driving conformational change. These results derived...
from GNM calculation agree overall with those reported by Tan et al., suggesting that for 17b antibody binding gp120 cores, the inner domain and bridging sheet are more flexible domains than the outer domain. However, results differ at the residue level. A rigid core for α-2 helix (residues 335–352) identified by FIRST is not identified by a specific GNM slow-mode minimum. Tan and Rader identified this core via comparison with CD4 unliganded gp120s (SIV and b12 bound), whereas our study focused on CD4 and CD4 mimic bound gp120s. They also observed that the flexibility index for the domains is dependent on either CD4 or CD4 mimic bound to gp120 with increased flexibility of the inner domain and decreased flexibility for the outer domain when scyllatoxin CD4 mimic is bound, as compared with when CD4 is bound, to gp120. Minor differences in slow-mode minima and maxima on a residue-level are delineated by the GNM method for GPO2 versus GPO3. A recent structure report of a CD4 unliganded YU2 gp120 reveals a less-structured bridging sheet where the β2/β3 and β21/β22 strands are dissociated from the interface of the inner and outer domain. The β21/β22 strands contain the longest continuous region of GNM slow-mode maxima observed in GPO1, GPO2, and GPO3. Moreover, all the regions of GNM slow-mode minima reported here do not exhibit any conformational changes in the unliganded YU2 gp120.

Analysis of communication propensities based on the elastic network model yielded hubs of residues which facilitate communication between residues, as reflected by their low commute times. Many of the good communicating pairs revealed with GNM are also observed in calculations from MD trajectories. Furthermore, few residues exhibit both GNM maxima and good communication propensities. In contrast, many residues that exhibit slow-mode minima also demonstrate good communication propensities and reside on β-strands of the outer domain. This implies, in the case of gp120, that the more structurally stable portions of the protein are equipped for more efficient communication. We also observe that positively correlated GNM fluctuations across large distances in the outer domain are facilitated by interleaving hubs of good communicators. These results are consistent with the correlated motions observed in four regions of gp120 using covariance web analysis from Essential Dynamics Calculations. This study broadly identified correlated motions in the β-bundle at the termini-proximal end of the inner domain; the gp120 bridging sheet domain, the V3 domain, a region composed of the LC (265–270), LD(278–283), LE (350–357), V5 (459–463), α-2 (334–348), and β-bundle (358–362, 374–379, 463–470) in the proximal end of the outer domain. These broadly defined regions overlap with specific communications hubs identified in Figure 6. The study of Liu et al., also describes the major relative motions of the domains in the unbound form of Gp120 namely, twisting of the inner domain and bridging sheet relative to the outer domain and a twisting of the outer domain relative to the inner domain, bridging sheet, and

Figure 7
Comparison of cross-correlations matrices. A. GNM residue fluctuations; B. GNM communication propensity from MD averaged snapshot; C. Cross-correlation of motions of C-α atoms around the averaged position from MD simulation.
V3 domain. This relative motion of the inner and outer domains correlates with the observation in this study that most of the regions of GNM slow-mode minimum occur in the outer domain along the interface of the inner domain.

We have elucidated the communication hubs in each gp120 domain. The present findings provide us with a topological basis for efficient mediators of communication in the gp120 molecule. The lack of interdomain communication in gp120 was at first, unexpected. However, communication propensity is delineated by the underlying protein structural network. Given the large rearrangement of the bridging sheet and inner domain revealed in recent structures, effective communication among the domains is unlikely. The two communicating hubs in the bridging sheet (residues 118–122 to 433–434) between the adjacent anti-parallel β-strands (β-2/β-3 and β-20/β-21) suggests that in the CD4 bound form, a signal may be transmitted by the association of these two previously unstructured regions. Indeed, steered MD studies indicate that the β-2/β-3 and β-20/β-21 are conformationally independent. The three communication hubs in the inner domain are not exclusively associated with GNM slow-mode minima. Nonetheless, two of these residues, V242 and T244, do exhibit GNM slow-mode maxima. The three communication hubs, 224–228, 242, 244–245, and 487–488 are located on the five stranded β-sandwich located in the N- and C-terminal portion of the inner domain. The molecular arrangement of the gp120 trimer as revealed by electron tomography indicates that the β-sandwich portion of the inner domain would project toward the center of the trimer axes and the interface of gp41. Recently, Pancera et al. describe the architecture of the inner domain as an invariant β-sandwich anchoring three mobile layers as exhibited in various CD4 bound and unbound forms of gp120 complexes. It would be premature to speculate that the three inner domain communication hubs may facilitate communication with other components of the viral spike, but they may have physiological significance.

Characterizing both GNM fluctuation and residue communication profiles may aid the interpretation of mutagenesis data. Residues that are characterized with GNM slow-mode minimum reflect the structural topology of the protein based on its underlying packing density. Thus, residues with GNM slow-mode minima would be fairly sensitive to mutations that alter the underlying protein topology. Comparing the consistency of GNM results with mutation data yields credence to the interpretation of the gp120 fluctuation profile revealed in this study. Many gp120 mutations that affect CD4 binding and have GNM slow-mode minima, such as T257, L259, N262, and Y384, suggesting good overlap between computational and biological data. Madani et al. have demonstrated that the NBD class of inhibitors has varying sensitivity to Phe43 cavity mutants (V255, T257, and S375), depending on the halogen substitution pattern on the NBD-phenyl ring. Both T257 and S375 display a GNM slow-mode minimum and good communication propensities as calculated from GNM and MD simulations. Mutation of T257 to either Ala or Ser negatively impacted relative viral enhancement compared with wild-type. Given the importance of T257 as a communication hub and its potential interaction with the NBD phenyl ring, it is not surprising that even the conservative change to serine would disrupt the communication network embedded in the topological fold in the cavity. Mutation of S375 to Gly reduced HIV sensitivity to enhancement by NBD compounds while the S375A mutation increased HIV sensitivity to all but the largest phenyl substituted analogues. This suggests that S375 and T257 serve as a fine-tuned sensor in the Phe43 cavity and that communication propensity and fluctuation minima are plausible explanations for the roles of these residues in inducing and/or stabilizing the CD4 bound conformation of gp120.

The site of antibody 17b and CCR5 coreceptor binding, overlap at the interface of the bridging sheet and outer-domain including the V3 loop (not included in this study). The coreceptor N-terminal and the second extracellular loop bind to gp120 at the bridging sheet and at the junction of the V3 loop and outer domain. GNM slow-mode minima and maxima do not map within the 17b and CCR5 coreceptor binding sites, as defined by the 1QAD crystal structure. However, communication hubs are exhibited at this coreceptor binding site. In the absence of a gp120-CD4-CCR5 crystal structure, a direct pathway for signal propagation between the Phe43 cavity and the chemokine receptor site could only be inferred. Yet, residues that may exhibit efficient communication propensities in the general chemokine receptor site binding region have been identified in this study. On the bridging sheet, the communicating network between β-2/β-3 and β-20/β-21 (residues 118–122 to 433–434) forms contacts with the antibody interface. Residues K121, R419, K421, and Q422 have also been shown to be important for CCR5 interactions. Residues, N295, H330, and N332, located in the outer domain near the base of the V3 loop are also posed for efficient communication. Analysis of the cross-correlation matrix of the C-α atomic motion from the average structure produced from the GPO3 MD trajectory confirms a positive correlation in motion between several of these residues (371 to 383–386 to 330–330 to 414–418) supporting the notion of signal propagation from the Phe43 cavity. Morra et al. have calculated communication propensity from Hsp90 ligand bound and unbound MD trajectories inferring signal propagation between distant domains. Furthermore, they also report that the correlation matrix from GNM is consistent with that obtained by MD and confirm that a signal propagation pathway extracted from GNM is similar to that inferred from MD. Certainly, further work on the gp120 system would entail...
MD derived communication propensities using the set of mutant gp120s with various ligands reported by Madani et al. as a basis for identifying a pathway for signal propagation.33 Experimentally, the dynamic and energetic coupling of the communication hubs could also be investigated by in silico alanine mutagenesis analysis66 where each residue is mutated to alanine successively, and the effect on the communication propensities and energetic coupling can be calculated, providing a measure of the contribution of that particular residue toward stability of the communication network.

A wealth of biological and structural data is available for the gp120/CD4/coreceptor complex elucidating structure function relationships. A new context for this information is provided by delineating points of minima and maxima in the fluctuation profile and translating topological features to communication propensities. Residues identified as GNM minima in the outer domain may represent a core folding nucleus found in the prebound form of gp120. With respect to the CD4 cavity, residues that are likely to be recognized in the unbound form and which may facilitate signal propagation have been identified (T257, L259, N262, Y384). Furthermore, regions of GNM maxima that contact the biphenyl in the cavity indicate flexible regions that may be amenable to accommodating a larger ligand. We hypothesize that minima in the inner domain may represent hinge regions, driving the structuring of the CD4 bound form of gp120. We conclude from GNM profiles that W427, a GNM maximum, surrounded by two residues, H105 and M475, of GNM minima, form a key structural element that stabilizes the formation of the CD4 cavity. Podesta et al. have demonstrated that the boundary between dynamic regions generalize to regions of catalytic activity in enzymes.67 On the opposite face of the Phe43 cavity from W427, are a number of GNM minima in the outer domain that comprise communication hubs which are topologically suited to transmit signal of cavity structuring to the chemokine receptor site. Our conclusions from this study of CD4 bound forms of gp120 are consistent with recently published gp120 structures revealing plastic regions of the inner domain and bridging sheet. GNM has enabled us to deduce differences in conformational dynamics, which might not be otherwise deciphered from the crystal structure of a CD4 bound form of gp120, and provided insights to protein–ligand interactions for the complex process of HIV gp120 envelope recognition, cell attachment, and viral entry.

### Materials and Methods

#### Protein modeling

Three X-ray crystal structures were prepared for normal mode and molecular dynamic calculations: CD4-bound HIV-1 gp120 core strain HXB2: PDB code 1G9N and the scyllatoxin mini-protein (CD4M47A) bound HIV-1 gp120 core strain YU2: PDB code 2I5Y. The deglycosylated core gp120 construction of envelop protein in the three crystal structures, has a 19- and a 52-amino acid residue deletions from both the N- and C-termini respectively, is devoid of the V3 variable loop and contains a tripeptide substitution (Gly-Ala-Gly) for the 67 residues of the V1/V2 loop. Residue numbering in the crystal structures and in this study conforms to the numbering used in the full-length gp120 sequence. A gap in the numbering exists from residues (127–194 and 296–330) corresponding to the substitution of the V1/V2 loops with the tripeptide and the deletion the V3 loops. For, 1G9M and 2I5Y, the missing V4 loop was added from the 1G9N crystal structure and minimized. Hydrogen atoms were added and tautomeric states and orientations of Asn, Gln, and His residues were determined with Molprobity (http://molprobity.biochem.duke.edu/).68,69 Hydrogens were added to crystallographic waters using MOE.70 The OPLS-AA force field in MOE was used, and all hydrogens were minimized to a root mean square (rms) gradient of 0.01, holding the heavy atoms fixed. A stepwise minimization followed for all atoms, using a quadratic force constant (100) to tether the atoms to their starting geometries; for each subsequent minimization, the force constant was reduced by a half until 0.25. This was followed by a final cycle of unrestrained minimization. Water molecules, Asn-linked acetyl-D-glucosamine, 2-(acetylamo)-2-deoxy-a-d-glucopyranose, and small molecules were removed before GNM and MD calculations.

#### Gaussian network model

In GNM, the protein is modeled as a network of residues, with each residue being represented by its α-carbon atom. Bonded and nonbonded pairs of residues located within an interaction cutoff distance $R_c$ (7.0 Å) are assumed to be connected by springs (or harmonic potentials) with a uniform spring constant $\gamma$, which is the single parameter (force constant) of the Hookean potential, proposed by Tirion.72 The topology of the structure is fully defined by the Kirchhoff matrix of inter-residue contacts, $\Gamma$, also known as the connectivity matrix, which in turn fully defines the equilibrium dynamics of the structure, or the most likely deformations near the folded state. For a network of $N$ residues, the elements of the Kirchhoff matrix $\Gamma$ are defined as

$$
\Gamma_{ij} = \begin{cases} 
-1 & \text{if } i \neq j \text{ and } r_{ij} \leq R_c \\
0 & \text{if } i \neq j \text{ and } r_{ij} > R_c \\
-\sum_{i \neq j} \Gamma_{ij} & \text{if } i = j 
\end{cases}
$$

The cross-correlations between the fluctuations $\Delta R_i$ and $\Delta R_j$ of the nodes $i$ and $j$ are given by55:

$$
\rho_{ij} = \frac{\langle \Delta R_i \Delta R_j \rangle - \langle \Delta R_i \rangle \langle \Delta R_j \rangle}{\sqrt{\langle \Delta R_i^2 \rangle \langle \Delta R_j^2 \rangle}}
$$
\[ \langle \Delta R_i \cdot \Delta R_j \rangle = \left( 3k_B T / \gamma \right) [\Gamma^{-1}]_{ij} \]  

(2)

where \( k_B \) is the Boltzmann constant, \( T \) is the absolute temperature, and \( \gamma \) is a uniform spring constant. The inverse of \( \Gamma \) is expressed in terms of the nonzero eigenvalues \( \lambda_k (1 \leq k \leq N-1) \) and corresponding eigenvectors \( u_k \) of \( \Gamma \) as:

\[ \Gamma^{-1} = \sum_{k=1}^{N-1} \lambda_k^{-1} u_k u_k^T \]  

(3)

which permits the mean square (ms) fluctuations of a given residue to be expressed as a sum over the contributions of all modes:

\[ \langle (\Delta R)_{ij}^2 \rangle = \sum_{k=1}^{N-1} \frac{3k_B T}{\gamma} (\lambda_k^{-1} u_k u_k^T)_{ii} \]  

(4)

where \( u_k \) and \( \lambda_k \) are the respective \( k \)-th eigenvector and eigenvalue of the Kirchhoff matrix, \( \Gamma \). The \( i \)-th element of \( u_k \) reflects the mobility of residue \( i \) in the \( k \)-th mode. \( \lambda_k \) scales with the frequency of mode \( k \), and \( \lambda_k^{-1} \) is a statistical weight, which suitably rescales the contribution of mode \( k \). Thus, the slowest mode has the largest contribution to the observed dynamics and the highest degree of cooperativity.55 The shapes of the slow-mode profile reveal the mechanism of the cooperative of global motions. The most constrained residues in these modes play critical mechanical role such as acting as hinge centers.44–48,55 The fastest modes, on the other hand, are localized to single residues that are usually tightly packed in the folded state. Being the most constrained residues, they are the latest to evolve (or reconfigure) and do not contribute significantly toward the global motion. In Eq. (4), the subscript \( ii \) designates the \( i \)-th diagonal element of the matrix enclosed in parenthesis. The X-ray crystallographic temperature factors (or B-factors) can be compared with the theoretical mean square (ms) fluctuations by the equation:

\[ B_i = \frac{8 \pi^2}{3} \langle (\Delta R_{i})^2 \rangle = \sum_{k=1}^{N-1} \frac{8 \pi^2 k_B T}{\gamma} (\lambda_k^{-1} u_k u_k^T)_{ii} \]  

(5)

**Communication propensities**

The residue fluctuations in GNM also determine the communication propensities between residues.51 An expression for commute time \( C(i,j) \) between residue \( i \) and \( j \), in terms of inverse of Kirchhoff matrix, \( \Gamma^{-1} \) is given by:

\[ C(i,j) = [\langle \Gamma^{-1} \rangle_{ii} + \langle \Gamma^{-1} \rangle_{jj} - 2\langle \Gamma^{-1} \rangle_{ij}] \sum_{k=1}^{N-1} \Sigma_{i \neq j} (d_k) \]  

(6)

which reduces to:

\[ C(i,j) = \langle \Delta R_i^T \Delta R_j \rangle \frac{[\gamma/3k_B T]}{\sum_{k=1}^{N-1} (d_k)} \]  

(7)

where the constants are as in Eq. (4), and \( d_k \) is the local interaction density at residue \( k \). The commute time between residues \( i \) and \( j \) is thus directly proportional to the fluctuation in the distance between these two residues, because the term in parentheses in the above equation is a constant for all pairs of residues. The mean square fluctuations play a dominant role in determining the communication propensities of a given pair of residues—larger the mean square fluctuation, longer the commute time.

**ACKNOWLEDGMENTS**

The authors appreciate the useful suggestions and discussions with Ivet Bahar in analyzing the GNM results and also thank Martha S. Head for suggesting this study.

**REFERENCES**


