Atomic insight into the CD4 binding-induced conformational changes in HIV-1 gp120[#]

Shang-Te D. Hsu and Alexandre M.J.J. Bonvin

Bijvoet Center for Biomolecular Research, Utrecht University, 3584CH Utrecht, The Netherlands

[#] Reproduced with permission from *Proteins: Structure, Function and Bioinformatics*, 2004, Vol. 55, 582-593

Abstract

The entry of HIV-1 into a target cell requires gp120 and receptor CD4 as well as co-receptor CCR5/CXCR4 recognition events associated with conformational changes of the involved proteins. The binding of CD4 to gp120 is the initiation step of the whole process involving structural rearrangements that are crucial for subsequent pathways. Despite the wealth of knowledge about the gp120/CD4 interactions, details of the conformational changes occurring at this stage remain elusive. We have performed molecular dynamics simulations in explicit solvent based on the gp120/CD4/CD4i crystal structure in conjunction with modeled V3 and V4 loops to gain insight into the dynamics of the binding process. Three differentiated interaction modes between CD4 and gp120 were found, which involve electrostatics, hydrogen bond and van der Waals networks. A "binding funnel" model is proposed based on the dynamical nature of the binding interface together with a CD4-attraction gradient centered in gp120 at the CD4-Phe43-binding cavity. Distinct dynamical behaviors of free and CD4-bound gp120 were monitored, which likely represent the ground and pre-fusogenic states, respectively. The transition between these states revealed concerted motions in gp120 leading to: *i*) loop contractions around the CD4-Phe43-insertion cavity; *ii*) stabilization of the four-stranded "bridging sheet" structure; and *iii*) translocation and clustering of the V3 loop and the bridging sheet leading to the formation of the co-receptor binding site. Our results provide new insight into the dynamic of the underlying molecular recognition mechanism that complements the biochemical and structural studies.

Introduction

The entry of Human Immunodeficiency Virus type-1 (HIV-1) into a target cell is an extremely intricate process (for reviews, see (Chan and Kim, 1998; Wyatt and Sodroski, 1998; Berger *et al.*, 1999; Poignard *et al.*, 2001)), which requires the recognition of the host cell receptor CD4 and co-receptors, mainly CCR5 and CXCR4, by the viral envelope glycoprotein (Env), consisting of gp120 and gp41. Env is organised into trimeric spike structures on the surface of the virion (Kwong *et al.*, 2000b) anchored in the membrane by the transmembrane subunit gp41 while gp120 is presented at the surface as the frontier for the recognition process. The initiation step involves binding of CD4 to gp120, which induces conformational changes in gp120 (Sattentau and Moore, 1991; Thali *et al.*, 1993; Moore and Binley, 1998) leading to the exposure and/or formation of antigenic epitopes recognised by the co-receptors (Trkola *et al.*, 1996; Wu *et al.*, 1996). Binding to the co-receptors takes place subsequently and gp41 undergoes conformational changes that mediate the fusion process by formation of the fusogenic structure bringing viral and target cell membranes into close vicinity. This complex mechanism involves a series of structural rearrangements in which conformational dynamics plays a crucial role.

Over the past decades, a large number of crucial elements involved in this system, in particular gp120, have been unraveled using a variety of approaches (Rizzuto *et al.*, 1998; Wyatt and Sodroski, 1998). The recent determination of the crystallographic structure of the core of gp120 in complex with CD4 and

the CD4-induced antibody (CD4i), which recognises the co-receptor binding site, provided tremendous structural insights (Kwong *et al.*, 1998; Kwong *et al.*, 2000a). Despite the highly variable sequence of HIV-1 gp120 throughout evolution, the topology of the core has been retained with some degree of sequence diversity (Kwong *et al.*, 1998; Kwong *et al.*, 2000a). It contains several disulphide bridges that buckle the flexible hypervariable loops to form knots at the periphery of the rigid core and basically define their boundaries. The heavy glycosylation at the "silent face" and the mobile variable loops together provide a shielding umbrella to evade immune system neutralization (Wyatt and Sodroski, 1998). CD4 is bound to gp120 predominantly *via* electrostatic interactions with a large, but mismatched interface. An unusually protuberant Phe43 of CD4 inserts into a receptive hole of gp120 (Kwong *et al.*, 1998) making up a "knob-and-socket" interaction (Moore and Binley, 1998). Evidence from a thermodynamics study has suggested that the binding of CD4 to gp120 induces substantial structural rearrangements primarily in the core structure, a truncated form of gp120 that was used in the crystallographic studies where the flexible N- and C-termini as well the variable loops, V1-V3, are absent (Myszka *et al.*, 2000) (see Figure 1). It is now generally accepted that the gp120/CD4 complex formation leads to the accessibility of the co-receptor binding site. The third variable (V3) loop, in particular, plays a central role here since it contains not only the principle neutralising determinant (PND) (Rusche *et al.*, 1988) but also the main co-receptor determinant (CXCR4 versus CCR5 usage) for HIV-1 at this stage (Berger *et al.*, 1999). X-ray crystallographic and Nuclear Magnetic Resonance (NMR) studies of various types of V3 loop fragments have shown a high degree of structural heterogeneity (Chandrasekhar *et al.*, 1991; Rini *et al.*,

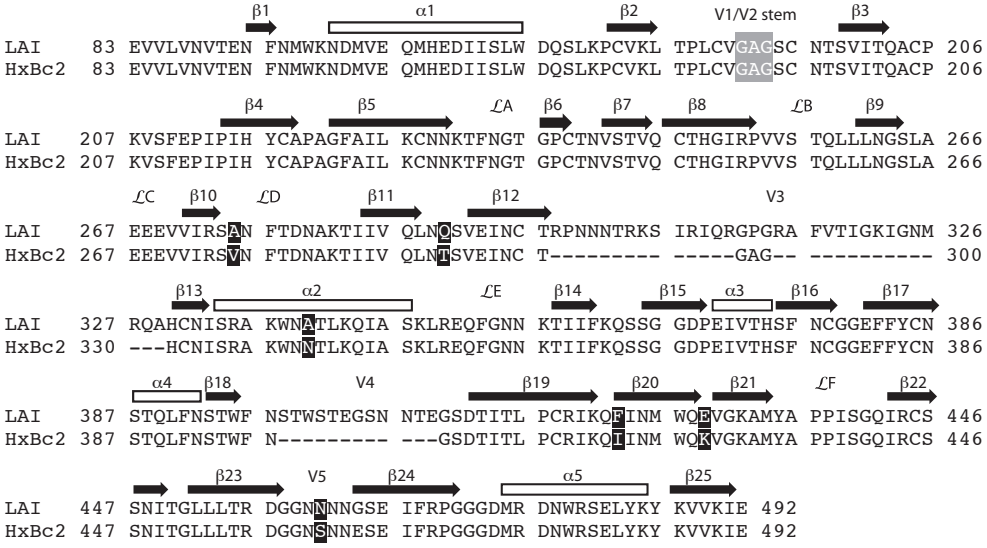

Figure 1. Primary sequence of the truncated form of LAI strain with the core structure and the V3 loop used in the current MD simulation with the numbering system according to the HXBc2 isolate. The secondary structure elements are numbered according to the X-ray core structure (PDB entry 1G9M, chain G, Kwong *et al.*, 1998) The truncated V3 loop and the disordered V4 loop that are absent in the coordinates of HXBc2 gp120 are indicated by dashes along the sequence of HXBc2. The truncated N- and C-termini are omitted for clarity. Black blocks depict differences in sequence composition of LAI versus HXBc2 that are present in the simulations.

1993; Ghiara *et al.*, 1997; Stanfield *et al.*, 1999; Tugarinov *et al.*, 1999; Tugarinov *et al.*, 2000; Sharon *et al.*, 2003), which may thus provide means for HIV-1 to evade neutralization by the immune system (Stanfield *et al.*, 1999). On one hand, biochemical studies have established the outline of the fusion process revealing its vast diversity and sophistication; on the other hand, results from the structural studies have provided static snapshots of the interaction between gp120 and its interacting partners at atomic resolution. The dynamics involved in this mechanism, however, remains elusive.

We have performed molecular dynamics (MD) simulations to investigate the conformational changes in gp120 induced upon complexation of CD4. By comparing the dynamical properties of gp120 in the free form and bound to CD4, we will show that CD4 binding substantially reduces the dynamical motions of gp120. An extensive intermolecular hydrogen bond network is formed at the interface and around the Phe43 binding cavity. We monitored motions around the CD4 binding site that are correlated to the V3 loop and trimerization interface, which likely represent the initial steps of the subsequent structural rearrangements required for co-receptor binding and gp41 mediated membrane fusion.

Material and Methods

Generation of the starting structures

The starting structure of gp120 with the truncated LAI strain sequence (SWISS-PROT accession number P03377) was obtained from SWISS-MODEL (Guex and Peitsch, 1997) using structural homologs as template (PDB entries 1G9M (Kwong *et al.*, 1998), 1G9N (Kwong *et al.*, 2000a) and 1GC1 for the gp120 core and 1CE4 for the V3 loop (Vranken *et al.*, 1995)). The particular strain was chosen because of ongoing collaborations with experimental groups (I. Braakman, Utrecht University and B. Berkhout, University of Amsterdam, personal communication). The primary sequence of gp120 of LAI differs from HxBc2 by only six amino acids in the core region (98.1% sequence identity, see Figure 1). Core gp120 is defined as the construct present in the crystal structure with the truncated V1-V3 and N- and C-termini. The V4 loop (residues 398-409), for which electron density is missing in all crystal structures, and the V3 loop, which has been proposed to undergo significant rearrangements upon CD4 binding, were modeled onto the g120 core, together with the six core mutations of the LAI strain, using the default settings of SWISS-MODEL (Guex and Peitsch, 1997). For the V3 loop, a NMR structure (residues 296-331; PDB ID: 1CE4) was taken for docking onto the core structure. This modelling and in particular the six point mutations are expected to have only minor effect on the overall structure of gp120. For example, the crystal structures of two HIV-1 variants, HxBc2 (1G9M) and YU2 (1G9N) with a sequence identity of 85.9%, show backbone (N, Cα and C') root-mean-square deviation (RMSD) of only 0.12 nm. The structural variation in core gp120 between LAI and HxBc2, which share 98.1% sequence identity, should thus be minimal and our starting model for the simulations can be considered accurate. For the simulation of the complex form (*Cplx*), the LAI gp120 model (residues 83-492) was superimposed onto the crystal structure of the complex (1G9M) to obtain the relative orientation with respect to the D1 domain of CD4 (CD4-D1, residues 1-99). Monomeric gp120 (*Free-gp120*) and CD4-D1 (*Free-CD4*) taken from the structure of the complex were used as starting structures in the free simulations for comparison.

The starting conformations of gp120 in the free and complex simulations were thus identical and correspond to the CD4-bound form.

Molecular dynamics simulation setup and structural analysis

The GROMACS 3.0 molecular dynamics package (Lindahl *et al.*, 2001) was used with the GROMOS 43A1 force field (Daura *et al.*, 1998). Starting structures of the free gp120 (*Free-gp120*), the free CD4 D1 domain (*Free-CD4*) and the gp120/CD4 D1 domain complex (*Cplx*) were individually solvated using the simple point charge (SPC) water model (Berendsen *et al.*, 1981) in rectangular periodic boxes with a 1.4 nm solute-wall minimum distance. After a first steepest descent energy minimisation with positional restraints on the solute, seven, three and ten chloride anions were introduced in *Free-gp120*, *Free-CD4* and *Cplx*, respectively, to obtain electro-neutralised systems. The resulting systems comprised 89608, 24054 and 95308 atoms for *Free-gp120*, *Free-CD4* and *Cplx*, respectively. A second energy minimisation was then performed, followed by five successive 20 picosecond (ps) MD runs with decreasing positional restraints force constants on the solutes (K_{posres} = 1000, 1000, 100, 10 and 0 kJ mol^{-1} nm^{-2}) prior to the 10 nanosecond (ns) production runs. A two femtosecond (fs) time step was used for the integration of the equations of motion. Solute, solvent and counter-ions were independently coupled to a reference temperature bath at 300K with a coupling constant τ_T of 0.1 ps (Berendsen *et al.*, 1984). The pressure was maintained by weakly coupling the system to an external pressure bath at one atmosphere with a coupling constant τ_p of 0.5 ps. Non-bonded interactions were calculated using twin range cutoffs of 0.8 and 1.4 nm. Long range electrostatic interactions beyond the cutoff were treated with the generalised reaction field model (Tironi *et al.*, 1995) using a dielectric constant of 54. The non-bonded pair list was updated every 10 steps. The LINCS algorithm (Hess *et al.*, 1997) was used for bond length constraining in conjunction with dummy atoms for the aromatic rings and amino group in side chains (Feenstra *et al.*, 1999). Owing to the large system sizes *ca* 2 ns were required to reach equilibrium. The analysis was performed therefore on the 2-10 ns segments for all systems.

Equilibrated trajectories (2-10 ns) of free and CD4-bound gp120 were merged for essential dynamics analysis (Amadei *et al.*, 1993). After diagonalising the backbone atom (N, Cα and C') mass-weighted covariance matrix the primary principle component (the first eigenmode) was extracted and identified as the transition mode between the two states. From a projection of the merged trajectory along this eigenmode the structures corresponding to the two extremes were obtained.

The β-turn conformations in the V3 loop were compared with several experimental structures of this loop (PDB entries: 1CE4 (Vranken *et al.*, 1995), 1B03 (Tugarinov *et al.*, 1999), 1QNZ (Tugarinov *et al.*, 2000), 1AI1 (Ghiara *et al.*, 1997), 1F58 and 2F58 (Stanfield *et al.*, 1999)). For all types of β-turns a minimum distance between the Cα(i) and Cα(i+3) of 0.7 nm was required. The criteria for backbone torsion φ/ψ angles for type II β-turn are: φ(i+1) = -60°, ψ(i+1) = 120°, φ(i+2) = 80° and ψ(i+2) = 0° and those for type VIII β-turn were: φ(i+1) = -60°, ψ(i+1) = -30°, φ(i+2) = -120° and ψ(i+2) = 120°, each torsion angle has a ±45° tolerance.

Results

In order to investigate the conformational changes in gp120 associated with CD4 binding, three molecular dynamics (MD) simulations of 10 ns production were performed in explicit solvent for the following systems: free gp120 (*Free-gp120*), the free D1 domain of CD4 (*Free-CD4*) and the gp120/CD4 D1 domain complex (*Cplx*). All systems showed stable trajectories with conserved secondary and tertiary structures (Table 1). Stability and equilibration of the systems was monitored by following the RMSD values as a function of time (Figure 2). The loop regions in free gp120, namely V1-V5, give rise to larger structural deviations than in the complex form (Figure 2 and Table 1). The larger RMSD values of the

Table 1. Structural statistics (2-10 ns)

Average backbone RMSD[a] with respect to the starting structure (nm)

	Gp120 all bb	Gp120 2° bb	CD4 all bb	CD4 2° bb	Gp120+CD4 all bb	Gp120+CD4 2° bb
Free	0.42(0.02)	0.16(0.02)	0.10(0.02)	0.08(0.01)		
Cplx	0.37(0.02)	0.14(0.01)	0.11(0.01)	0.09(0.01)	0.40(0.02)	0.35(0.02)

Secondary structure elements statistics[b] (number of residues)

	β sheet	α helix	Turn
Gp120 initial[c]	102	52	17
Free	96(7)	56(3)	12(4)
Cplx	111(7)	50(2)	13(3)
CD4 initial[c]	41	0	9
Free	47(3)	5(1)	11(3)
Cplx	47(4)	6(2)	12(3)

a. The backbone (bb) positional RMSD values were calculated with respect to the starting structure (based on PDB entry 1G9M (Kwong *et al.*, 1998), see Materials and Methods) after superposition on the respective secondary structure elements (2°) as defined by DSSP (http://www.sander.ebi.ac.uk/dssp/). The high RMSD values of gp120+CD4 in the complex are due to relative motions of the two proteins while separate analysis of the individual molecule shows small deviations from the starting crystal structure. The larger deviations of the overall backbone in gp120 is due to the high mobility of the V3 loop. Standard deviations are indicated in parentheses.

b. The secondary structure content was calculated with DSSP excluding the modelled V3 and V4 loops.

c. The total numbers of residues of gp120 and CD4 in our simulations are 346 and 99, respectively (*cf.* Material and Methods).

entire gp120 backbone in the free form primarily stem from the re-positioning of the loops near the CD4 binding site as can be readily seen by comparing the snapshots of the trajectories of gp120 in the free and CD4 bound forms (Figure 3). The snapshots also illustrate that the V3 loop, which started from an identical initial configuration in the free and complex simulations, evolves into distinct configurations with well-separated spatial distributions in the two simulations. In addition, substantial reduction in the amplitude of loop motions, in particular in the V1/V2 stem, can also be observed (this will be discussed in detail in the following section). Given the compactness of CD4-D1, its RMSD values from the crystal structure are, in both runs, small and constant regardless of the presence of gp120 (see Table 1). Despite the different dynamical behaviour of gp120's flexible loops, which is reflected both in RMSD values and per

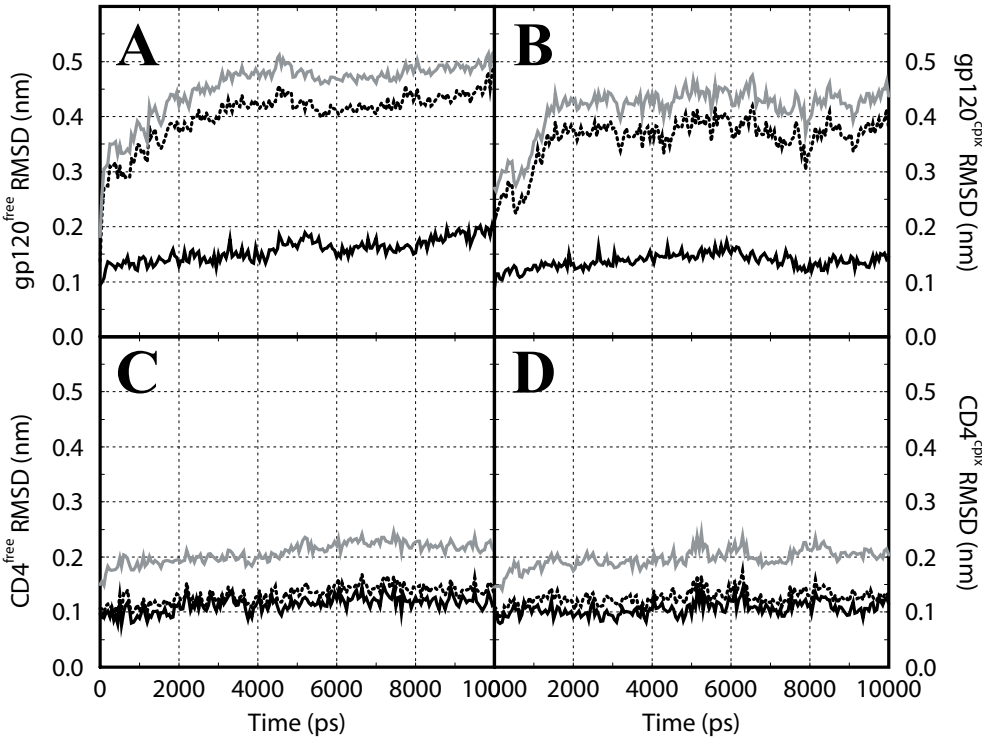

Figure 2. Time evolution of the RMSD from the starting structure during the 10 ns simulations of gp120 and CD4. **A.** *Free-gp120*; **B.** gp120 in *Cplx*; **C.** *Free-CD4*; **D.** CD4-D1 in *Cplx*. The backbone (N, Cα and C') RMSD values of the secondary structure elements defined in the crystal structures and the complete backbone are shown in solid and dashed lines, respectively. The all atom (including hydrogen atoms) RMSD values are shown in grey lines. The RMSD values were calculated every 2 ps after superimposition on the backbone of the secondary structure element and are plotted as a running average over a 50 ps window.

residue fluctuations, the internal energies of both gp120 and CD4 are not much perturbed upon binding (Table 2). Also worth noting is that the energy of binding ($E_{complex} - E_{free\ gp120} - E_{free\ CD4}$) amounts to –130 kJ/mol (considering only protein-protein and protein-solvent interactions and thus assuming that no significant changes are taking place in the bulk solvent and counter ions). This is in fair agreement with the change in enthalpy upon complexation of gp120/CD4 (-62±2 kJ/mol at 310K) reported by Doyle and co-workers (Myszka *et al.*, 2000).

Substantial reduction of gp120 loop mobility and conformational change upon CD4 binding

A clear difference in mobility is found between the variable loops of gp120, V1-V5, and the conserved core region. Binding of CD4 substantially decreases the motions of the loops interacting with CD4, especially the V1/V2 stem and the V5 loop. It also induces a remarkable repositioning of the V3 loop that remains highly mobile in both states (Figure 3). Analysis of the backbone RMS fluctuations, expressed in terms of crystallographic temperature factor (B-factor), as a function of residue sequence for *Free-gp120*

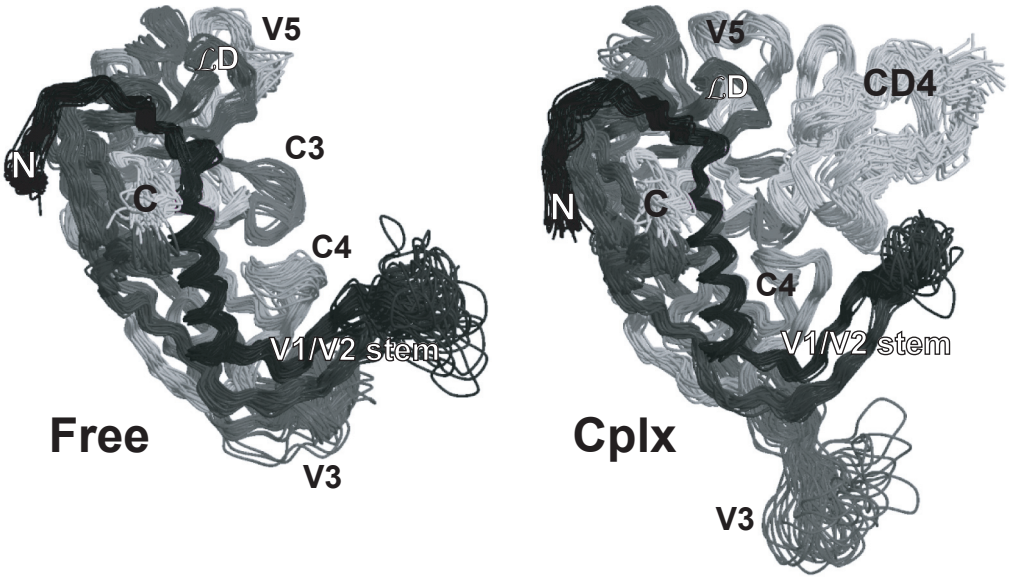

Figure 3. Snapshots of the MD simulations. **Left:** Free gp120; **Right:** Gp120 in complex with CD4 D1 domain. The backbone Cα traces are taken every 200 ps from the MD trajectories between 2-10 ns. The structures are superimposed on the secondary structure backbones of gp120. Gp120 is coloured from black at the N-terminus to light grey at the C-terminus. CD4-D1 structure ensemble from the *Cplx* simulation is shown in white. Loops that showed significant differences in conformation and/or flexibility are indicated (C3 consists of β15 and α3; C4 consists of β20 and β21, which form the four-stranded bridging sheet with β2 and β3 in the V1/V2 stem). Figures were generated using Molscript (Kraulis, 1991) and Raster3D (Merritt and Bacon, 1997).

and *Cplx* (Figure 4) reveals that most of the high-mobility sites (except V3 and V4) coincide with the CD4 binding-site. The calculated B-factors of gp120 match well with the crystallographic data with the exception of the V1/V2 stem. In CD4, however, they are significantly smaller than the experimental ones. Discrepancies in the overall B-factors amplitude are also present between the two crystal structures of free CD4, 1CDH (Ryu *et al.*, 1990) and 3CD4 (Wang *et al.*, 1990), where the latter shows similar sequential patterns but 30-40% larger overall values. Although a quantitative comparison between the simulated and experimental B-factors may not be applicable due to the difference between experimental and simulation conditions (Hünenberger *et al.*, 1995), it is apparent that the observed reduction in loop mobility upon binding near Phe43 in CD4-D1 is consistent with the experimental findings (Figure 4 inset).

The sizeable reduction of mobility upon binding is primarily determined by the intermolecular contacts in the crystal structure, *e.g.* the V1/V2 stem, ℒD and V5 regions as well as the C3 and C4 regions that contact the protruding CD4-Phe43 (Kwong *et al.*, 1998). The attenuation of the loop motions arises, to a large extent, from the formation of many intermolecular hydrogen bonds, most of which are also present in the crystal structure (Table 3, Figures 4 and 5). The large loop motions in gp120 observed in both the free and complex simulations are consistent with the fact that the crystallisation of gp120 was only possible when the variable loops, *i.e.* V1-V3, were truncated (Figure 1); the introduction of CD4 was necessary to further stabilise the ternary complex. Despite its shorter

Table 2. Statistics of non-bonded energies[a] (2-10 ns)

Coulomb's electrostatic energy (kJ/mol)		
Free gp120	Gp120 internal	-38262 (363)
	Gp120-solvent	-30428 (674)
	Total	-68690 (766)
Free CD4	CD4 internal	-12298 (314)
	CD4-solvent	-11761 (497)
	Total	-24059 (588)
Gp120/CD4 complex	Gp120 internal	-37923 (416)
	Gp120-solvent	-29930 (699)
	CD4 internal	-12765 (289)
	CD4-solvent	-11295 (503)
	CD4-gp120	-1030 (177)
	Total	-92943 (1015)
Lennard-Jones (van der Waals) energy (kJ/mol)		
Free gp120	Gp120 internal	-12430 (115)
	Gp120-solvent	-2795 (144)
	Total	-15225 (184)
Free CD4	CD4 internal	-3406 (51)
	CD4-solvent	-671 (78)
	Total	-4077 (93)
Gp120/CD4 complex	Gp120 internal	-12276 (103)
	Gp120-solvent	-2635 (137)
	CD4 internal	-3395 (50)
	CD4-solvent	-477 (25)
	CD4-gp120	-452 (76)
	Total	-19235 (196)

a. The non-bonded energies were calculated with the GROMOS96 force field (Daura *et al.*, 1998) using a twin range cutoff of 0.8 and 1.4 nm with a reaction field correction (see Material and Methods). The energies are the sum of short- (SR) and long range (LR) terms; 1-4 terms were not included. Standard deviations are indicated in parentheses.

loop length, resulting in higher structural restriction, the V4 loop was still disordered and unresolved in the crystal structures (Kwong *et al.*, 1998; Kwong *et al.*, 2000a), probably because of its intrinsic conformational heterogeneity and remoteness from the CD4 binding interface.

Intermolecular hydrogen bond network and Phe43 cavity

The increase in rigidity observed in the complex originates primarily from the formation of intermolecular hydrogen bonds. Although many of them are already present in the crystal structure, some are only formed during the molecular dynamics simulation. Our simulation allows, in addition, to assess the dynamical nature of these critical interactions. This can be done by analysing their occurrence throughout the trajectory. Many long-lived intermolecular hydrogen bonds bridge the tips of the mobile loops of gp120 interacting with CD4 and a hemisphere of CD4-D1 (Figure 5). This hydrogen bond network nicely encloses the docking cavity of Phe43, which is the key element for the binding mechanism. A

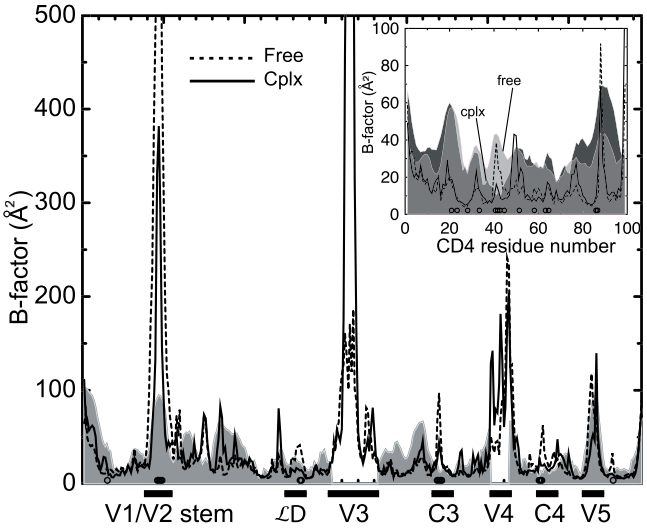

Figure 4. Comparison of gp120 Cα atoms temperature factors (B-factors) calculated from the RMS fluctuations in the 2-10 ns simulations of *Free-gp120* (dashed lines) and *Cplx* (solid lines). The B-factors for CD4-D1 are shown in insets at a different scale. Experimental B-factors taken from the ternary gp120/CD4/CD4i crystal structure (PDB entry 1G9M, Kwong *et al.*, 1998) are shown in grey area. The B-factors of CD4-D1 in the free form (light grey area overlaid onto the complex one) are taken from the free CD4 structure (PDB entry 3CD4)(Wang *et al.*, 1990). They are comparable to the B-factors in the complex except for the Phe43 binding site Note that the V3 loop was truncated and the electron density of the V4 loop was missing. Residues forming stable intermolecular hydrogen bonds in the complex form are indicated with open circles along the horizontal axes (for definitions, see Table 3).

large number of side chain-side chain or side chain-backbone hydrogen bonds cover a large area of the complex interface in a non-specific manner. For instance, the side chain hydroxyl group of S42 in CD4 is able to form several hydrogen bonds by hopping amongst various backbone hydrogen bond acceptors in gp120 that are in its close proximity (Table 3). In addition, there are a few stable backbone-backbone hydrogen bonds in the vicinity of the Phe43 cavity (Table 3) involving residues of gp120 (S365, G366, G367 and D368) that are well conserved amongst various HIV-1 strains. The importance of these residues has been highlighted by structure-based mutagenic studies in which mutations of S365, G366 and D368 in gp120 diminished the affinity of the CD4-binding site specific antibody b12 (Saphire *et al.*, 2001). Overall, the residues involved in these intermolecular hydrogen bonds gave rise to an average intermolecular Coulomb's electrostatic energy of –624±124 kJ/mol, accounting for 61% of the total intermolecular Coulomb's electrostatic energy (Table 2).

These observations allow us to propose a functional role for the hydrogen bond network: the non-specific, loose, flexible and wide-spread hydrogen bonds and salt-bridges and the specific, tight, rigid and confined ones generate an affinity gradient that drives the insertion of the Phe43 phenyl ring into its binding cavity in gp120. Once inserted, it is locked in its "knob-and-socket" geometry by the specific hydrogen bond network in the proximity of the binding pocket. The hydrophobic contacts between the CD4 Phe43 side chain and the gp120 cavity then provide the short-range stabilising factor. These

Table 3. Intermolecular hydrogen bond and salt bridge statistics[a] (2-10 ns)

Intermolecular hydrogen bond				
CD4		Gp120		Occurrence (%)
Backbone - backbone				
F43	N	G473	O	11.5
L44	O	D368	N	67.8
K46	N	S365	O	14.2
K46	N	G366	O	15.7
Side chain – backbone				
K35	Nζ	A281	O	15.9
S42	Oγ	Q428	O	69.9
S42	Oγ	E429	O	61.5
S42	Oγ	E429	N	31.2
L44	N	D368	Oδ	92.6
D53	Oδ2	G367	N	11.1
R59	Nη1	E429	O	45.9
R59	Nη1	V430	O	55.2
Q64	Nε2	C126	O	92.4
Q64	Nε2	V127	O	18.0
Q64	Nε2	A129[b]	O	29.3
N66	Nδ2	A129[b]	O	25.8
Side chain – side chain				
K22	Nζ	E429	Cδ	77.8
Q25	Nε2	D474	Oδ	93.8
K29	Nζ	N280	Oδ	83.6
R59	Nε	D368	Oδ	80.7
Q64	Nε2	S131	Oγ	15.2
E87	Cδ	K97	Nζ	61.1
D88	Cγ	K97	Nζ	14.8

a. Intermolecular hydrogen bonds and salt bridges (side chain-side chain hydrogen bonds) occurring for more than 10% during the 2-10 ns period are listed. A hydrogen bond is considered to exist when the donor-hydrogen-acceptor angle is larger than 120º and the donor-acceptor distance is smaller than 0.28 nm.

b. A129 is one of the linker residues GAG that were used in the crystal construct (1G9M) to replace the V1/V2 loops (128-194) (Kwong *et al.*, 1998).

contacts are extremely well maintained during the simulation bearing very small structural variations (Table 4); translation and rotation of the phenyl ring with respect to the cavity were limited within 0.2 nm and 60º, respectively (data not shown).

CD4 binding induced bridging sheet stabilisation

The so-called "bridging sheet" is a four-stranded β-sheet that consists of β2 and β3 in the V1/V2 stem and β20 and β21 in the C3 region. It interposes between the inner and outer domains of gp120 and is thought to be flexible and disordered in the absence of CD4. Recent thermodynamic studies on the binding of gp120 and CD4 have demonstrated unprecedented large entropy changes, which were

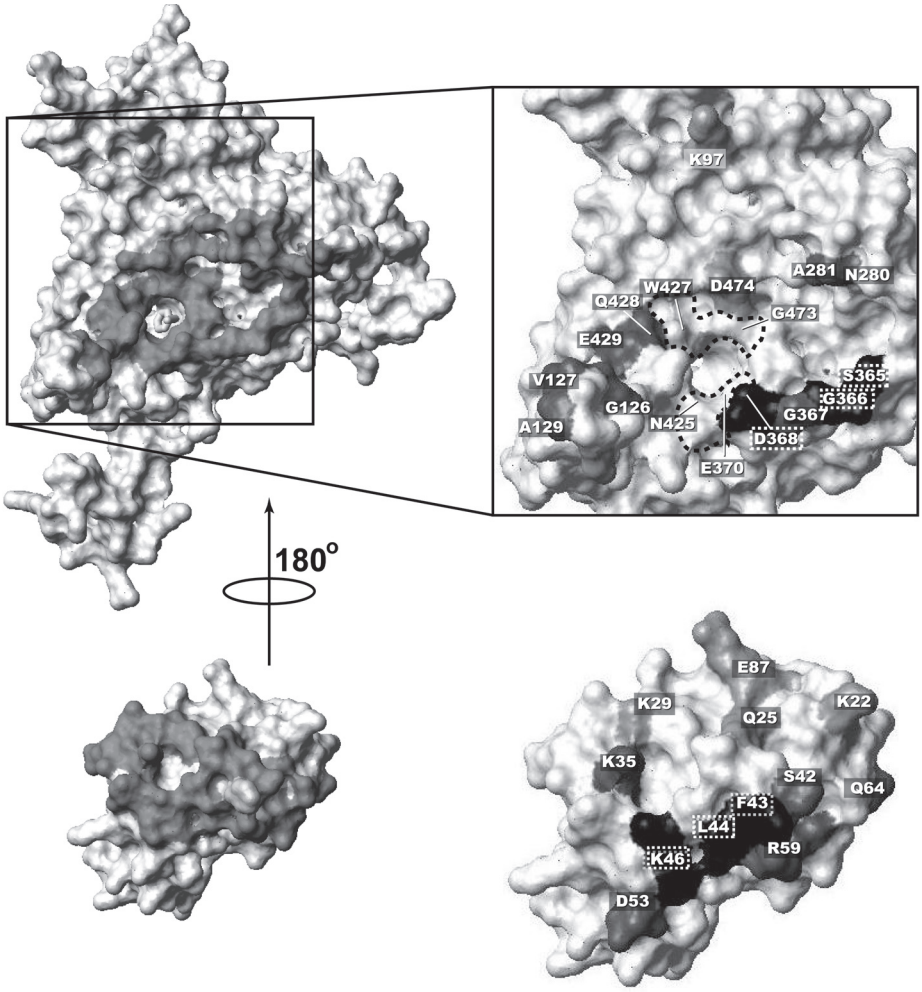

Figure 5. Mapping of the intermolecular hydrogen bond network onto the intermolecular contact interfaces. **Left panel:** The contact van der Waals surfaces of gp120 and CD4-D1. Heavy atoms of gp120 and CD4-D1 that have intermolecular distances within 0.5 nm in the starting structure are coloured dark grey. The side chain of CD4-Phe43 is shown in light grey and as sticks in gp120 to indicate the binding cavity. **Right panel:** Residues involved in intermolecular hydrogen bonds. The types of hydrogen bonding is colour-coded as: *light grey*: side chain-side chain hydrogen bond; *dark grey*: side chain-side chain and/or side chain-backbone hydrogen bond; *black*: backbone-backbone and/or side chain-backbone hydrogen bonds (for details, see Table 3). Residues that form the receptive CD4-Phe43 cavity (E370, N425, W427 and G473) are outlined with dashed lines. Note that several residues that do not have intermolecular contacts in the starting structure form intermolecular hydrogen bonds during the molecular dynamics simulations (gp120: K97, V127 and A129; CD4: K22, D53 and E87). The figures were generated using MolMol (Koradi *et al.*, 1996).

Table 4. Statistics of CD4 Phe43-gp120 cavity native contacts (%)[a] (2-10 ns)

Residue	Atom	Occurrence
E370	Cγ	100.0
E370	Cδ	99.5
N425	Cβ	94.1
N425	C	94.1
N425	O	99.0
W427	N	99.1
W427	Cα	100.0
W427	Cβ	83.2
G473	Cα	99.2
G473	C	98.9
G473	O	100.0

a. Eleven heavy atoms of gp120 have interatomic distances with any atom of the CD4 Phe43 side chain within 0.35 nm in the crystal structure (1G9M, Kwong *et al.*, 1998) A native contact during the simulation is considered to exist if the distance between the listed atoms and the centre of the Phe43 phenyl ring is less than 0.6 nm.

proposed to be possibly accounted for by a stabilisation of the bridging sheet and a restriction of the interdomain motions upon CD4 binding. Detailed structural and dynamical information is, however, still lacking.

The conformational changes of the bridging sheet upon removal of CD4 were monitored *in silico* by RMSD matrix analysis (Figure 6 lower right panel). The bridging sheet undergoes a series of transient structural changes with a major one at about 8 ns, after which it adopts a conformation somewhat similar to the ones in the 4.5-5.5 ns segment. The same analysis on the simulation of the complex showed little change throughout the 10 ns simulation (Figure 6 upper left panel). Several interstrand backbone hydrogen bonds between the antiparallel β2 and β21 that delimit the inner and outer domains of gp120 also show different hydrogen-acceptor distance distribution in the two simulations: While the total numbers of the interstrand backbone hydrogen bonds throughout the simulations are the same, for *Free-gp120*, the hydrogen bond length distribution peaks at 0.188 nm with a half width of 0.048 nm whereas for *Cplx* it peaks at 0.182 nm with a narrower half width of 0.035 nm. The shorter hydrogen bond distances in the complex are indicative of an increased stability of the bridging sheet upon CD4 binding while it remains flexible otherwise. This is consistent with the model proposed in the thermodynamic studies mentioned previously.

Conformational changes and polymorphism of the V3 loop upon CD4 binding

The V3 loop has been targeted as a prominent candidate to tackle the viral fusion problem. Based on structural (Kwong *et al.*, 1998), mutagenic (Rizzuto *et al.*, 1998) and antigenic analyses (Wyatt *et al.*, 1998), co-receptor binding was proposed to require the V3 loop in concert with the bridging sheet. These form the co-receptor binding site that is created only upon CD4 binding. A large panel of monoclonal antibodies (MAbs) based on V3 loop fragments has been elicited against several primary isolates from different HIV-1 clades (subtypes) (Gorny *et al.*, 1993; Nyambi *et al.*, 1998; Gorny *et al.*, 2002). Several V3 loop-derived peptides have also been devised to inhibit viral entry into target cell in

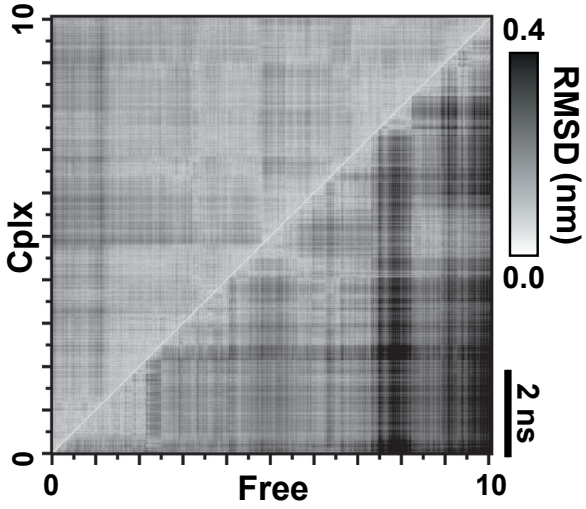

Figure 6. Pair-wise backbone RMSD matrix of the bridging sheet in the free (lower right panel) and the complex form (upper left panel). The four β-strands are defined as β2: 109-113; β3: 198-202; β20: 423-427; β21: 430-434. Each dot represents a positional RMSD between two conformations taken from the respective trajectories indicated on the axes and is colour-coded accordingly to the scale shown on the right. The RMSD values were calculated on backbone atoms of the bridging sheet (β2, β3, β20 and β21) after superimposition on the backbone atoms of β2 and β3. Conformations were taken every 10 ps. An equilibrated conformational sampling period is found when an off-diagonal region shows a continuous low RMSD (white to light grey).

a co-receptor-specific manner (Sakaida *et al.*, 1998; Verrier *et al.*, 1999; Basmaciogullari *et al.*, 2002). Successful vaccine design has however been hindered by the underlying structural polymorphism of the V3 loop. The well-conserved region, GPGR, can indeed adopt various type of β-turns and is found in various configurations when bound to different MAbs (Rini *et al.*, 1993; Ghiara *et al.*, 1994; Ghiara *et al.*, 1997; Stanfield *et al.*, 1999; Tugarinov *et al.*, 1999; Tugarinov *et al.*, 2000; Sharon *et al.*, 2003). Note that previous structural studies were only performed on V3 loop fragments and little is thus known about its conformation when attached to the core structure. We therefore investigated the dynamics and conformation of the V3 loop both in free gp120 and in the CD4-bound form.

As mentioned above the V3 loop is highly mobile and occupies an ample space in the vicinity of the co-receptor binding site. Different localisations in the free and CD4-bound states can be clearly distinguished in the corresponding trajectories (Figure 3). Relocation of the V3 loop, upon CD4 binding, towards the basal part of the bridging sheet agrees with the model proposed by Sodroski and co-workers based on their mutagenesis study on the co-receptor CCR5 binding (Rizzuto *et al.*, 1998). In their comprehensive study, the removal of the V3 loop (residues 298-329) was found to abolish the co-receptor CCR5 binding completely. Many residues (K121, T123, K207, L317, E381, F383, I420, K421, Q422, P438, R440 and G441) were also identified to result in more than 90% loss in CCR5 binding when mutated. Comparing the solvent accessible surface area (SAS) of these residues in the free and complex simulations reveals an increase of SAS in the complex form only for I420 (15%) and Q422 (31%) while a small decrease is observed for R440 (8%). Other residues are

either unaffected or buried in both the free and complex forms by the V3 loop. The changes in SAS of the above three residues amount, however, to a gain of *ca* 1.10 nm^2, which may be an indication of the proposed exposure of the binding site. The clustering of the V3 loop and the bridging sheet can also be inferred from the lowering of the electrostatic Coulomb s energy between the V3 loop as a whole (residues 298-329) and the subset of residues that are important for CCR5 binding: -438±60 kJ/mol in the complex form versus -345±101 kJ/mol in the free form. The abundant basic residues of the V3 loop (six arginines and two lysines, see Figure 1) are thought to facilitate the recruitment of the generally acidic chemokine co-receptor subsequent to the binding of CD4 (Dragic *et al.*, 1998; Farzan *et al.*, 1998; Rabut *et al.*, 1998). They generate a positive electrostatic potential that occupies an enormous space centred at the basal region of the V3 loop. Moving from the free to the complex form relocates the positive potential toward the co-receptor binding site while neglecting the V3 loop drastically reduces this positive electrostatic potential (Figure 7)

Structurally, the V3 loop shows a high propensity of β-hairpin formation in our simulations but the position of the turn is poorly defined (Figure 8). The conserved GPGR sequence shows several transient β-turns, including type II and type VIII β-turns in the two simulations. Other tight turns were found at RGPR, QRGP and IQRG, as a result of one to three residue upstream shifts. Comparison with the available GPGR structures (see Material and Methods) gives backbone RMSD

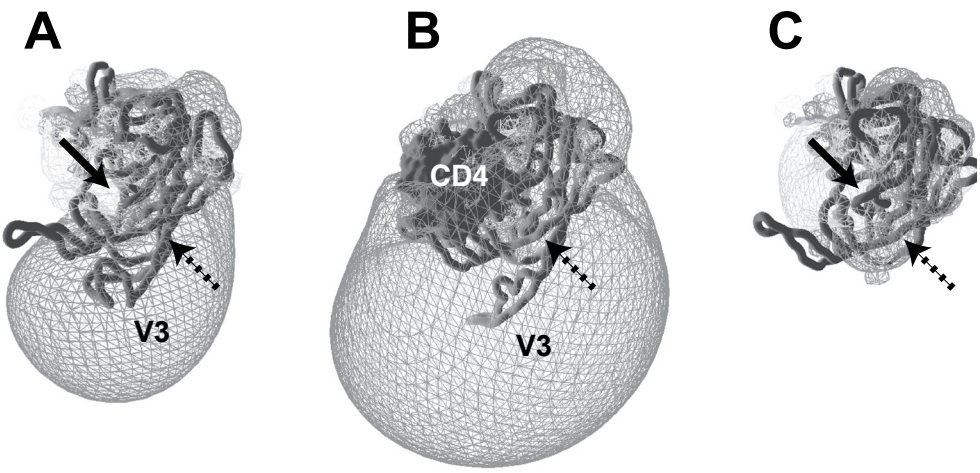

Figure 7. Electrostatic potential surface of gp120. **A.** Free gp120. **B** Gp120 in complex with CD4-D1. CD4-D1 is labelled and is shown in van der Waals surface. **C.** Gp120 in the complex form without CD4-D1 and the V3 loop. The core structure of gp120 is identical to that in **B** and CD4-D1 and the V3 loop were manually removed from the complex structure. The light and dark grey meshes correspond to electrostatic potentials of –2 and +2 kT/e, respectively. Negative electrostatic potentials are primarily present at the CD4 binding site of gp120. They were calculated using GRASP (Nicholls *et al.*, 1991) with protein and solvent dielectric constants of 2 and 80, respectively. The ionic strength was set to zero. The structures are taken from the snapshots at 6 ns both in the free and complex trajectories, which represent the most populated configuration of the V3 loop based on RMSD matrix analysis. The CD4 and CCR5 binding sites are indicated by solid and dashed arrows, respectively.

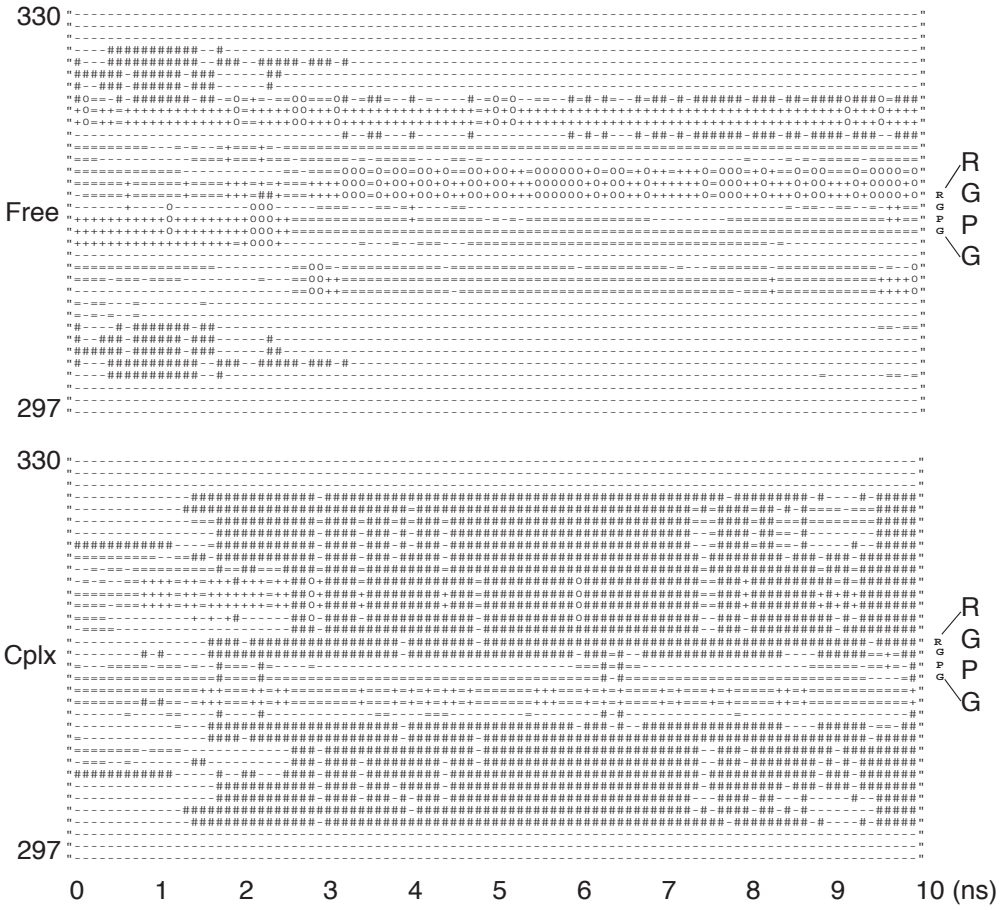

Figure 8. Secondary structure evolutions of the V3 loop (residues 297-330) as a function of the simulation time. Secondary structure elements are coded as following: Coil (-), β-sheet or β-bridge (#), bend (=), turn (+) and α- or 3_{10}-helix (0). Each block corresponds to a time step of 100 ps. The position of the conserved residues, GPGR, is indicated at the right end of each diagram.

values between 0.2 and 0.6 nm with the various antibody-bound structures irrespective of the effect of CD4 complexation (data not shown).

Essential dynamics analysis reveals concerted loop motions in gp120 upon CD4 binding

The complexity of molecular motions observed in MD simulations can be simplified by essential dynamics analysis (Amadei *et al.*, 1993), whereby sets of essential and correlated motions are extracted. Such an analysis allows distinguishing high frequency local motions, which are restrained and harmonic in nature and contain less structural information, from "essential", large amplitude global motions. In order to focus on the motions associated with CD4 binding such an analysis was performed on merged trajectories from the free and complex forms of gp120 and of CD4-D1. Two distinct distributions along the first

eigenvector, *i.e.*, with the largest eigenvalue, can be found, which correspond to the well separated free and complex states of gp120 (Figure 9A & B). Already the second eigenvector of the two trajectories shows overlap in positional fluctuation displacement suggesting similar motions between the two states. Narrow Gaussian displacement distributions appear after the first few eigenvectors suggesting restrained harmonic motions common to the two states. Projections of the two extremes of the first eigenvector along the merged trajectory on the average structure illustrate the major difference between the two states (Figure 9C). Note that the intrapolation between the two extremes does not represent the transition pathway but merely highlights the structural differences.

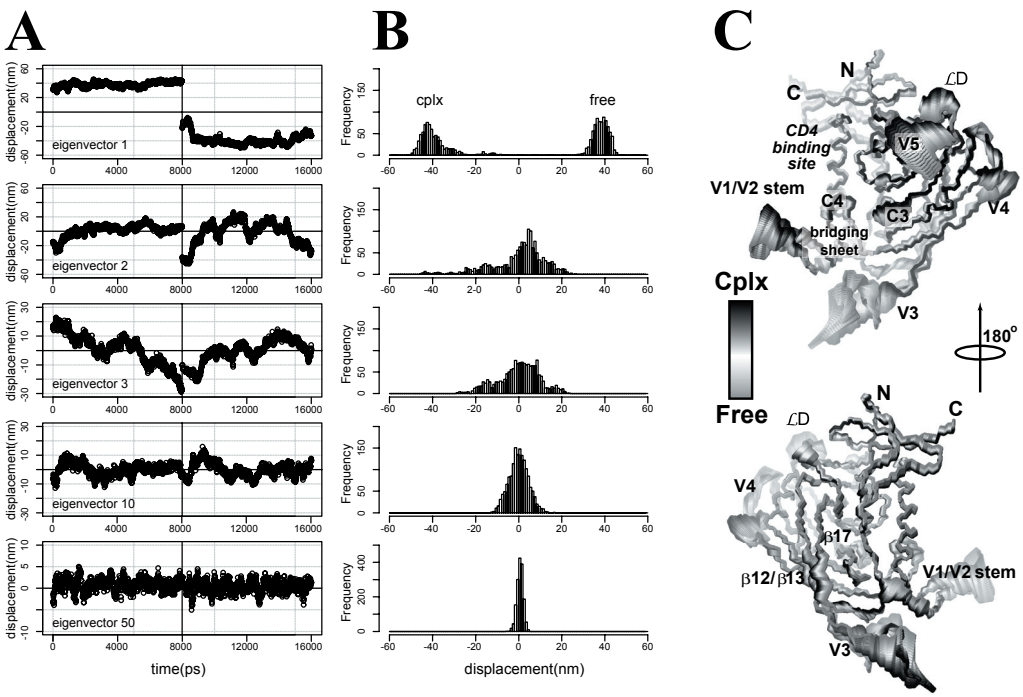

Figure 9. CD4 binding-induced conformational changes of gp120 extracted *via* essential dynamics analysis. **A.** Projections of the total atomic fluctuations of the merged trajectory (free: 0-8 ns and complex: 8-16 ns) along the selected eigenvectors. The eigenvectors are numbered with respect to the amplitude of displacement in descending order. **B.** Histograms of the distributions of the displacement of each eigenvector. Distinct distribution can only be found in the first eigenvector. Note that the scale is different for the 50th eigenvector (bottom row). **C.** Projections of the two extremes of the first eigenvector along the merged trajectory onto the average structure. The linear intrapolation between the two extremes is coloured from light grey (free) to black (complex) with white colour in the midst of the two to highlight the primary structural differences between the two states. The C3 and C4 regions of gp120 that are part of the Phe43 cavity show significant closure of their lid upon CD4 binding. The V3 loop shows the largest conformational change. The view from the back side of the structure after 180° rotation (*bottom*) reveals propagated structural perturbations at the interior of gp120, involving β17, which is adjacent to the Phe43 cavity as well as the two β-strands, β12 and β13 that connect the V3 loop.

In gp120, a correlated loop-contraction primarily around the CD4 binding site is revealed with, in particular, a pronounced closure of the lids of the Phe43 cavity (Figure 9C). This correlated motion increases the curvature of the CD4 binding site leading to a gain in interfacial complementarity. The concerted contraction involves primarily the V1/V2 stem, V5 and, to a less extent, the C3 and C4 regions. These finger-like structures, in particular the V1/V2 stem and V5, are in a much more open and relaxed state in the free form. The most prominent transition occurred in the V3 loop, which is driven by CD4 binding towards the basal region of the bridging sheet. In contrast to the substantial conformational changes in gp120 no major structural transition upon binding was identified for CD4-D1 suggesting that the principle modes of motion are not affected by complexation and that only their amplitudes are reduced.

Discussion

We have shown by MD simulations that CD4 binding reduces the mobility of various loops of gp120 around the binding site and induces conformational changes that effectively lead to the wrapping up of a hemisphere of CD4-D1. While substantial changes occur in gp120, in CD4-D1, only the Phe43 containing loop showed a slight reduction of mobility upon complexation. The intermolecular interactions can be categorised into three levels: non-specific side chain-side chain or side chain-backbone hydrogen bonds (or salt-bridges), specific backbone-backbone hydrogen bonds and hydrophobic contacts between both partners resulting in the insertion of CD4 Phe43 into the receptive cavity in gp120. The large-amplitude motions of the V1/V2 stem and V3 and V4 loops may provide a shielding umbrella that masks the CD4 binding site in the *monomeric* form of gp120 as proposed by Kwong *et al.* based on their thermodynamic finding (Kwong *et al.*, 2002). The observed entropy loss upon binding of CD4 to gp120 (Myszka *et al.*, 2000; Kwong *et al.*, 2002) is a clear indication of a reduction of flexibility and is in line with our observations. When CD4 is attracted into the vicinity of the binding site, the mobility of the V1/V2 stem, V5 loop and £D reduces. The extensive intermolecular hydrogen bond formation and van der Waals contacts at the gp120/CD4 interface result in the closure of C3 and C4 regions leading to the formation of the required Phe43 binding cavity. They also reduce the overall mobility in the regions involved in the intermolecular contacts between both partners.

The formation of the extensive intermolecular hydrogen bonds as well as the stabilisation of the bridging sheet may account for the experimentally observed entropic loss (Myszka *et al.*, 2000; Kwong *et al.*, 2002), which is compensated by a gain in enthalpy through the intermolecular interactions described here. In contrast, the V3 loop remains flexible after its relocation induced by the binding of CD4 and no entropy cost is thus paid for this process. These observations are in agreement with the experimental findings that the main entropic changes occur within the core structure of gp120 (Myszka *et al.*, 2000).

One can summarise the above observations into a "binding funnel" model where dynamics and different interaction modes are coupled leading to efficient recognition and specific affinity (Tsai *et al.*, 1999). During the search of receptor CD4, gp120 is constantly subject to immune system attack. A defined binding pocket based on rigid body docking would require exhausting geometry search while antibodies neutralisation *via* structural epitope recognition is also taking place. Gp120 devised, therefore, a multi-level binding mechanism as such that the conformation with the specific binding cavity is embedded in an ensemble of less selective conformations facilitating target search: *i)* the charged residues located in the

highly mobile loops generate long range electrostatic attraction that guide CD4 to its binding interface consisting of a non-specific hydrogen bond network bearing less structural selectivity and energetic trapping; *ii)* the loose side chain-side chain hydrogen bonding that confines gp120 and CD4 into close proximity leads to the formation of specific backbone-backbone hydrogen bonds and subtle changes in the loops around CD4 lead to the closure of the lid of the receptive cavity; *iii)* while the specific hydrogen bonds are formed, the Phe43 phenyl ring plugs into its binding cavity. Only one molecule that perfectly matches with the last two steps, CD4, will be able to accomplish the structural rearrangements that weld the complex structure together, thereby completing the first stage of the fusion process: the gp120/CD4 recognition. It should be noted that we did not simulate the binding process and therefore no kinetic or free energy data could be obtained from the current study. Therefore, our "binding funnel" model cannot be interpreted in term of a free energy surface of binding; it only presents a *geometrical* and *chronological* view of the binding process.

The ability of anti-CD4-binding-site MAbs to neutralise HIV-1 has been associated with their affinity for the trimeric envelope virion structure, rather than monomeric gp120 (Poignard *et al.*, 2001). Recent thermodynamics study also suggests that the monomeric gp120 may act as decoys to the immune system (Kwong *et al.*, 2002). Most antibodies but b12 and 4KG5 are in fact elicited against the flexible monomeric gp120 with suboptimal angle of approach to the CD4 binding site such that they are sterically hindered by the interaction between the variable loops in gp120 (Zwick *et al.*, 2003). The ability to bind oligomeric gp120 is in fact the determinant to neutralise the virus (Roben *et al.*, 1994; Sattentau and Moore, 1995). On the one hand, the interpretation of our observed conformational changes is confounded by the limited knowledge about the trimeric state of gp120, the monomeric gp120 crystal structure being the only structural template available to date. On the other hand, CD4 is the unique element for the initiation of the CD4-dependent viral entry pathway. The formation of the gp120/CD4 complex is apparently the common ground for recognition, regardless of the oligomeric state. This actually demonstrates the remarkable plasticity of gp120 where the quaternary constraint imposed by the trimerisation of the envelope proteins may well be generated *via* the binding to CD4 alone, assuming that the structure of the monomeric gp120/CD4 complex resembles that in the oligomeric state. Although not presented here, we do observe correlated motions upon CD4 binding on the "back-side" of gp120, which has been proposed to be involved in the trimerisation process, either directly or more likely *via* interactions with the transmembrane trimerisation domain gp41. Under the assumption of an unique gp120/CD4 complex structure, the gp120/CD4/CD4i crystal structure has provided invaluable information about the gp120/CD4 interface, which has led to a large number of structural and biochemical studies such as structural mapping by monoclonal antibodies (Saphire *et al.*, 2001; Xiang *et al.*, 2002). Yet, it still lacks the dynamical information regarding the conformational changes occurring in gp120 upon CD4 binding, the limiting factor to the subsequent viral entry process. Some intermolecular hydrogen bonds, mostly mobile side chain-side chain hydrogen bonding (Figure 5), were only identified in our simulation. The geometry and rigidity of the well-defined CD4-Phe43 cavity, which is composed of some mobile residues in the C3 and C4 regions, is only induced upon CD4-binding (Figure 3). Furthermore, formation of the complementary gp120/CD4 interface requires correlated loop contractions as revealed by essential dynamics analysis. This process also leads to a stabilisation of the bridging sheet while the two β-hairpins forming it (the V1/V2 stem and C3) are somewhat parted

in the free form. Correlated motions at the putative trimerisation interface are also observed. We can only speculate that the perturbation at the trimerization interface, propagated from the gp120/CD4 interface, may trigger the release of tension stored within the coiled-coil structure of gp41 as proposed in the "spring-loaded" model (Eckert and Kim, 2001).

Conclusion

Our simulations of free and CD4-bound gp120 represent two distinct states, namely the relaxed ground state and the contracted, excited or pre-fusogenic state of gp120, respectively. Large amplitude loop motions in gp120 are attenuated and show concerted contraction upon CD4 binding. The associated entropic cost is compensated by the wealth of intermolecular interactions at the CD4 interface ranging from non-specific to specific hydrogen bonds and hydrophobic contacts around and at the Phe43 binding cavity, in agreement with the large enthalpy/entropy compensation measured experimentally. This differentiated mode of interaction from long range electrostatic attraction *via* non-specific and specific short-range interactions, in combination with the dynamical nature of the system, allowed us to propose a binding funnel model for the gp120/CD4 interaction. In addition, the complexation also drives clustering of the V3 loop and the bridging sheet that generates a highly positive electrostatic attraction gradient for subsequent co-receptor binding. In line with the proposed model deduced from the mutagenesis mapping of the co-receptor binding site (Rizzuto *et al.*, 1998), our results provide a plausible explanation for the functional relationship between the V3 loop and CD4-binding and the otherwise inefficient CD4-independent entry pathway (reviewed by (Berger *et al.*, (1999)). We have here demonstrated the sophisticated plasticity of gp120, from a rigid core to a floppy exterior shielding that provides recognition specificity without compromising the capacity to evade attack from the immune system.

Acknowledgements

The authors thank Prof. Ineke Braakman and Eelco van Anken (Department of Bio-Organic Chemistry, Bijvoet Center, Utrecht University, The Netherlands) and Prof. Ben Berkhout (Department of Human Retrovirology, Academic Medical Center, University of Amsterdam, The Netherlands) for careful reading and helpful discussions. This work was sponsored by the National Computing Facilities (NCF) for the use of supercomputer facilities, with financial support from the Netherlands Organization for Scientific Research (NWO).