Structure of an unliganded simian immunodeficiency virus gp120 core

Bing Chen[1], Erik M. Vogan[1,2], Haiyun Gong[1], John J. Skehel[3], Don C. Wiley[1,2]* & Stephen C. Harrison[1,2]

[1]*Children's Hospital Laboratory of Molecular Medicine, Harvard Medical School, and* [2]*Howard Hughes Medical Institute, 320 Longwood Avenue, Boston, Massachusetts 02115, USA*
[3]*National Institute of Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK*

* Deceased

Envelope glycoproteins of human and simian immunodeficiency virus (HIV and SIV) undergo a series of conformational changes when they interact with receptor (CD4) and co-receptor on the surface of a potential host cell, leading ultimately to fusion of viral and cellular membranes. Structures of fragments of gp120 and gp41 from the envelope protein are known, in conformations corresponding to their post-attachment and postfusion states, respectively. We report the crystal structure, at 4 Å resolution, of a fully glycosylated SIV gp120 core, in a conformation representing its prefusion state, before interaction with CD4. Parts of the protein have a markedly different organization than they do in the CD4-bound state. Comparison of the unliganded and CD4-bound structures leads to a model for events that accompany receptor engagement of an envelope glycoprotein trimer. The two conformations of gp120 also present distinct antigenic surfaces. We identify the binding site for a compound that inhibits viral entry.

The envelope glycoproteins of HIV and SIV are the molecular agents of cell attachment and membrane fusion[1]. They are trimeric assemblies of gp160 polypeptide chains, which are cleaved during transport to the surface of an infected cell into two fragments known as gp120 and gp41 (refs 2–4). Cleavage enables the protein to undergo a series of conformational changes when it encounters the receptor for these viruses, CD4, and their co-receptor, CXCR4 or CCR5, on the surface of a suitable host cell[5–8]. The first of these changes, probably confined largely to gp120, accompanies receptor binding[9,10]. It stabilizes a conformation with which a co-receptor can then associate[7,8,11]. Co-receptor binding may induce further changes, leading to dissociation of gp120 from the membrane-anchored gp41. The latter then refolds through a series of steps that lead ultimately to fusion of viral and target-cell membranes[12,13].

The structures of gp120 and gp41 at the end of this sequence of conformational changes have been determined in a series of X-ray crystallographic studies[12–14]. Analysis of the postfusion gp41 structure has been particularly important for deriving a picture of the fusion process and for understanding the mechanism of peptide fusion inhibitors[13,15,16]. The structure of CD4-bound, HIV-1 gp120, in complex with a monoclonal Fab that recognizes the co-receptor site, has provided a framework for analysing envelope antigenicity[14,17]. The unliganded, prefusion structure of gp120 and the structure of the prefusion trimer (including the prefusion conformation of gp41) have resisted high-resolution analysis.

The structure of SIV gp120 in an unliganded conformation, described here, now fills one of those lacunae. As in the studies of the CD4-bound structure[14], we have used the gp120 'core', from which two large loops of highly variable sequence, V1–V2 and V3, as well as amino- and carboxy-terminal segments, have been deleted (Fig. 1a). Comparison of the new structure of unliganded gp120 with that of the CD4-bound protein shows that part of the molecule—the 'inner domain' (see below)—undergoes unexpectedly extensive conformational rearrangement upon receptor binding. In the process, coherent sites assemble for CD4 and co-receptor interaction. Because the protein we have crystallized is fully glycosylated, we can visualize directly the extent to which oligosaccharides coat its molecular surface. Knowledge of the unliganded conformation allows us to model how the protein might appear in a gp120/gp41 trimer and to picture the overall conformational change induced by contact with receptor.

Crystallization of the unliganded SIV gp120 core

Design of the fragment that we crystallized followed closely the strategies that led to crystallization of CD4-bound HIV-1 gp120 (refs 14, 18, 19). The gp120 cores of SIVmac 32H (ref. 20) and HIV-1 HXBc2 have 35% sequence identity and over 70% sequence similarity; alignment of the two sequences is unambiguous. Both have seven disulphide bonds in corresponding positions. Of the 13 glycosylation sites on the SIV protein and 18 on the HIV-1 protein, nine are conserved or shifted by no more than one or two residues. We substituted short linkers (GAG) for the V1–V2 and V3 loops and deleted 43 and 22 residues from the N and C termini, respectively. We obtained crystals as described in the Methods (see also ref. 21). In contrast to reported experience with the HIV-1 gp120 core[22], deglycosylation prevented rather than facilitated crystallization. The crystals were small; we could record diffraction only to spacings of about 4 Å. The structure determination (see Methods and ref. 21) relied on two isomorphous derivatives, density averaging among three non-isomorphous crystal forms, and use of SeMet substituted protein to locate methionine residues.

Description of the structure

Overview

High overall sequence conservation between SIV and HIV gp120, their common receptor and co-receptors, and the properties of the SIV gp120 core structure described in this paper all support our interpretation, that most of the differences between our structure and that of CD4-bound HIV-1 gp120 reflect conformational changes induced by receptor binding rather than differences between the HIV-1 and SIV proteins. The unliganded gp120 core has the bipartite character first described for the CD4-bound form (Fig. 1), with inner and outer domains designated in relation to the location of N and C termini. The most extensive structural changes are in the inner domain, for which the sequence is even more conserved than for the outer domain, suggesting that a shared pattern of required conformational changes has constrained sequence variation. Four conserved inner-domain disulphide bonds lock in structural elements that must move with respect to

each other; the more globular outer domain has only three disulphides. Moreover, the conformational differences we observe are all consistent with predictions from studies of CD4 binding to gp120 in solution. For example, residues that make up the chemokine-receptor site, which requires CD4 binding to form[7,8,23], are not adjacent in our structure, whereas residues at which mutations generate resistance to a group of entry-inhibiting drugs all face a well-configured pocket[24]. Finally, the protein we have crystallized binds tightly to a panel of monoclonal antibodies raised against fusion-competent SIV gp120/gp41 trimer (Supplementary Fig. S1). We therefore refer to the structure reported here as the 'unliganded gp120 core', to distinguish it from the 'CD4-bound gp120 core' previously described[14]. In contexts in which we believe it important to remind the reader that one derives from SIVmac 32H and the

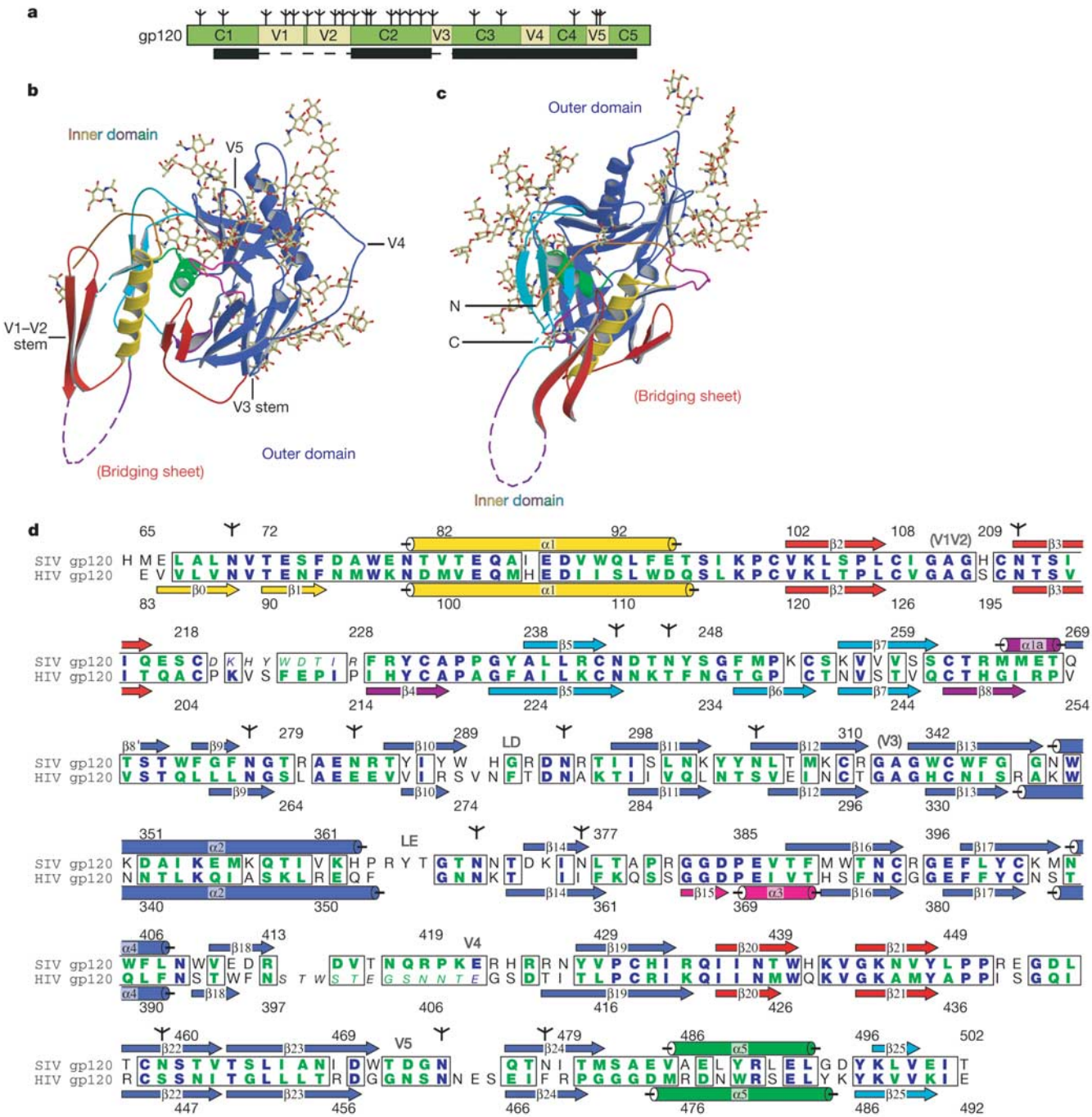

Figure 1 Structure and sequence of SIV gp120 core. **a**, Diagram of sequence elements in gp120 and definition of its core. The branched symbols mark glycosylation sites. **b**, Overall structure of the unliganded SIV gp120 core, with polypeptide chains as ribbon diagram and carbohydrates as stick models. Outer domain is in blue; inner domain, coloured according to substructure (N terminus, orange; α1, yellow; three-strand sheet, cyan; outer/inner domain transition, purple; α5, green). The four strands that form the bridging sheet in the CD4-bound conformation are in red and labelled as 'bridging sheet' in parentheses. Disordered residues are shown as dashed lines. The stumps of truncated variable loops V1, V2 and V3 are indicated and the intact variable loops V4 and V5 are also labelled. **c**, The same as **b**, rotated by 90° about a vertical axis. **d**, Alignment of the gp120 core sequences of SIVmac 32H and HIV-1 HXBc2. Residues in blue are identical; those in green are conservative substitutions; disordered residues are in italic. Glycosylation sites are indicated by tree-like symbols. Secondary structures of both proteins are shown as arrows for β-strands and rods for α-helices. They are coloured following the scheme used for the structure.

other from HIV-1 HXBc2, we add the SIV or HIV-1 designation.

The outer domain has, with some important local exceptions, essentially the same structure in the two states. The inner domain has reorganized markedly. Indeed, the inner domain turns out not to be a single, coherent structure that can shift as a rigid body, but rather a collection of distinct substructures that move independently with respect to each other when the unliganded and CD4-bound structures are compared. These substructures (colour-coded in the figures) include: a four-turn α-helix (α1); a β-ribbon (the 'V1–V2 stem', half of the bridging sheet in the CD4-bound state; β2–β3); a 3-strand β-sheet, with two successive strands (β5 and β7) augmented by the C-terminal strand (β25); and a short α-helix (α5) at the outer-domain/inner-domain junction. The N-terminal segment, imperfectly defined at a variable crystal contact, shifts approximately together with the 3-strand sheet. As predicted by Kwong and co-workers[14], the bridging sheet is absent in the unliganded state of gp120. Each of its two β-ribbons is ordered, but a space of 20–25 Å intervenes between them.

The kernel of the outer domain, against which CD4 docks, is an 8-strand, antiparallel β-barrel. The distal end of the barrel extends into a 6-strand β-sheet, which cradles a four-turn α-helix. Some differences between SIV gp120 and HIV-1 gp120 are in the lengths of loops in this domain. Thus, the V4 and V5 loops are slightly shorter in the SIV protein, and loop LE, just C-terminal to the helix and even more variable than V5, is slightly longer (Fig. 2). Two differences between our structure and that of CD4-bound gp120 are clearly owing to CD4 interactions. One is in the connection between a strand of the distal sheet (β14) and a strand (β16) of the barrel: this 'CD4-binding loop', with a conserved GGDPE sequence, moves when CD4 associates and presents an extended strand (β15) for main-chain hydrogen bonding with CD4's C″ ridge. The other is in the orientation of the loop (β20–β21) that forms one of the two β-ribbons of the bridging sheet (Fig. 2d).

Glycans

Secreted proteins from insect cells generally have mannose-rich (rather than complex) oligosaccharides. We find density for sugars at all 13 glycosylation sites of the SIV gp120 core. Seven of the glycans are clearly α(1–6) fucosylated, and eight are ordered at least out to the mannose branch site. The longer oligosaccharides form 'glycan clusters' on the molecular surface, probably through hydrogen-bonding networks. A catalogue of the glycan structures is included in ref. 21.

Variable loops

The gp120 core contains neither the V1–V2 nor the V3 loop, but we can infer their approximate locations from the positions of stumps left by their truncation. In the unliganded conformation, their stumps project in almost opposite directions (Figs 1a and 2b), and even allowing for uncertainties about their possible conformations (in SIVmac 32H, the deleted part of V1–V2 contains 99 residues, three disulphide bonds and seven N-linked glycans, which may form a compact structure decorated with sugars, and V3 contains 27 residues and 1 N-linked glycan), it seems unlikely that they make direct contact within a monomer. The V4 loop, disordered in the CD4-bound HIV-1 HXBc2 gp120 core structure[14], lies at a lattice contact in our crystals. It adopts an open conformation that extends away from the body of the outer domain. The amino-acid residues in this loop are almost entirely polar, and we expect that it might be more mobile when not constrained by crystal packing. Unlike the HIV-1 V4 loop, it contains no glycosylation sites. It thus creates an exposed ridge of protein surface above a sea of glycan. Its prominence may explain why neutralizing anti-SIV antibodies often recognize V4 (see refs in Supplementary Fig. S1).

Inner domain: comparison with CD4-bound structure

There are two useful ways to compare the unliganded and CD4-bound structures. If we choose a common orientation for the nearly

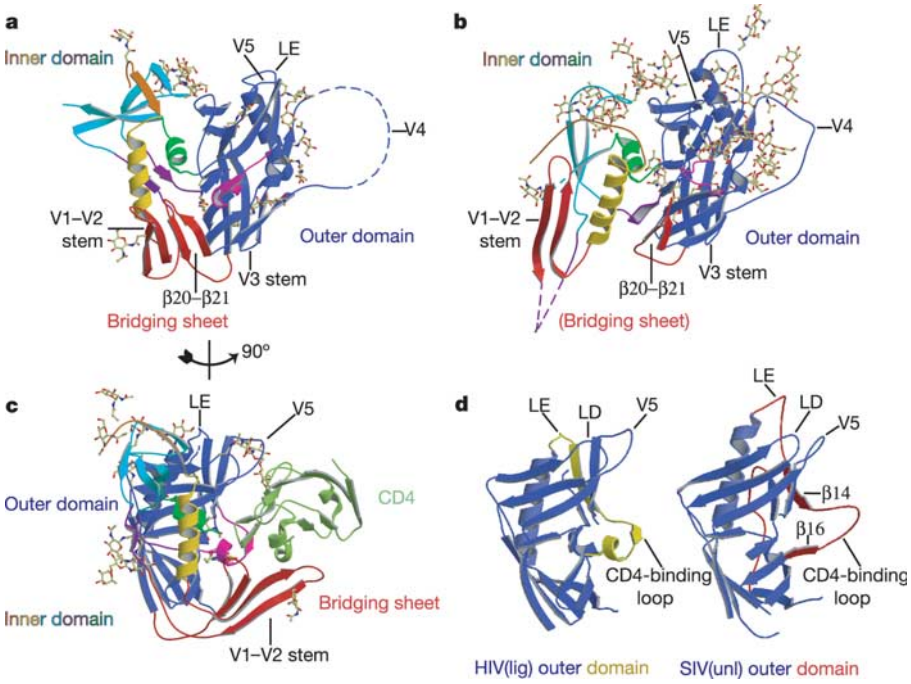

Figure 2 Comparison of glycosylated, unliganded SIV gp120 core with deglycosylated, liganded HIV-1 HXBc2 gp120 core[14]. **a**, Conformation of deglycosylated HIV gp120 core, complexed with two-domain CD4 and the Fab fragment of human neutralizing monoclonal 17b, with the outer domain on the right and the inner domain on the left. CD4 and Fab have been omitted for clarity. Twelve disordered V4-loop residues are shown as a dashed line. Colours as in Fig. 1b. **b**, SIV gp120 core structure, in the same view and colours as in panel **a**. **c**, Interaction of gp120 core and CD4, shown by a side view (90° from the view in panel **a**) of HIV gp120 core complexed with CD4 (light green). **d**, Comparison of outer domains of liganded (lig) HIV gp120 core (left) and unliganded (unl) SIV gp120 core (right). The differences are highlighted in yellow and red, and the CD4-binding loops are labelled. LD and LE are loop designations, following the convention of ref. 14.

invariant outer domains, as in Fig. 3a, b, the displacements and rotations of the inner-domain and bridging-sheet components are readily apparent. When CD4 binds, the three-strand antiparallel sheet rotates by 30°, the bridging sheet forms, and the four-turn α-helix (α1) shifts away from the outer domain. Connecting segments, such as the link between the V1–V2 stem and the three-strand sheet or the short α-helix between the outer domain and the C-terminal strand, follow in apparent response to the movements of the secondary-structural elements. The approximate centre-of-mass displacements of inner-domain substructures (with outer domains superposed) are 28 Å, 15 Å, 12 Å and 13 Å, for the N-terminal segment, the helix α1, the V1–V2 stem, and the β5–β7–β25 sheet, respectively.

The α1 helix is amphipathic. In the unliganded conformation it lies between the two separated β-ribbons of the bridging sheet, with its polar face exposed between them (Fig. 3a). In the CD4-bound conformation, it rotates to expose its polar face distal to the bridging sheet, which closes up over it (Fig. 3b). Intermediate positions of α1 would bury a set of charged residues, and intermediate conformations would therefore be unstable. This observation supports our inference, that the conformation we see represents a distinct, unliganded state, resembling the conformation present on unliganded Env trimer, rather than one of a continuum of flexible conformations selected by the crystallization conditions.

If we imagine that CD4 binding causes gp120 to shift relative to gp41, but that the latter remains essentially unaltered during this first of the series of conformational events leading to fusion, then the best frame of reference for relating the conformational change in the gp120 core to the gp120/gp41 trimer is one in which the N and C termini of the core and the three-strand inner-domain sheet from which they emerge are fixed and all the other structural elements in gp120, including the outer domain, move with respect to them. In this frame, we have the picture shown in Fig. 3c, d. Relative to the inner-domain sheet, the outer domain rotates outwards and 'upwards' by 35°. The bridging sheet closes up, with a nearly 40 Å displacement of the V1–V2 stem; the site for co-receptor interaction that it forms now faces away from the assumed direction of gp41 and therefore towards the target cell.

Receptor and co-receptor binding sites

Neither the receptor (CD4) nor the co-receptor (CXCR4 or CCR5) site is properly formed in the unliganded conformation of the gp120 core (Fig. 4a). In the unliganded conformation, the CD4-binding loop projects away from the centre of the outer domain, and the β20–β21 ribbon, one half of the bridging sheet, tucks in beneath it (Figs 1b and 3a). Together, the α-helices of the inner domain, the CD4-binding loop, and the β20–β21 ribbon create a long, narrow cavity, lined principally with hydrophobic side chains (Fig. 4b). We identify it as a binding site for small molecules that inhibit HIV-1 entry (see below). CD4 interacts with a face of the outer domain that is partly concealed within this cavity. In the conformation stabilized by CD4 binding, the bridging sheet can close up to create the co-receptor binding surface, which is flanked by the V1–V2 and V3 loops (Figs 3b and 4a). Co-receptor specificity depends on V3, which could influence binding either directly by interacting with co-receptor or indirectly by covering part of the bridging sheet.

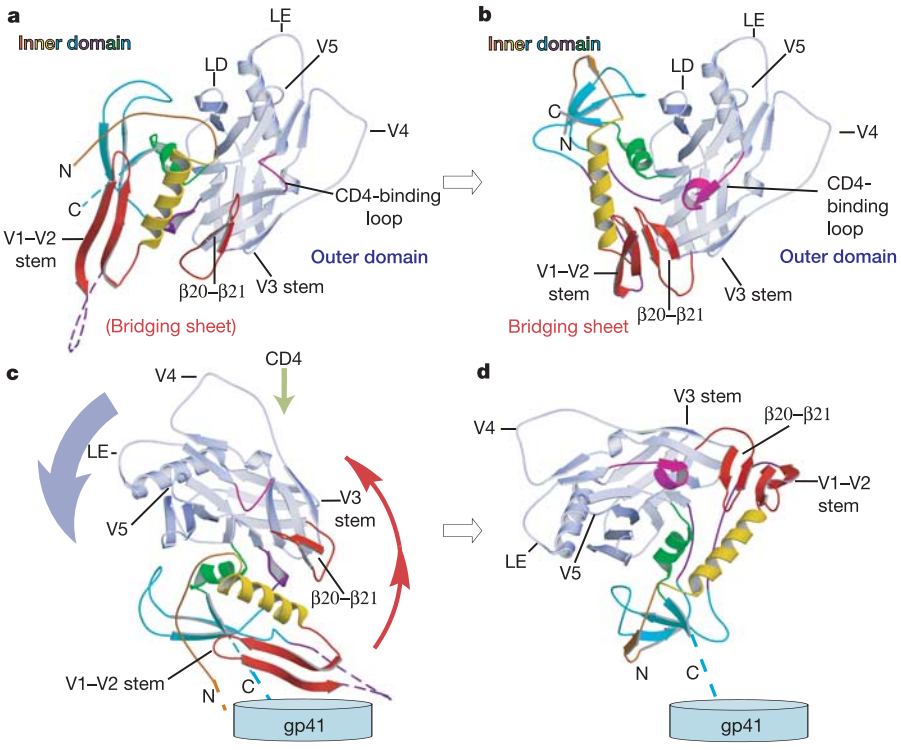

Figure 3 Conformational changes induced by CD4 binding. **a**, Unliganded SIV gp120 core structure, shown in the same colour scheme as in previous figures. The colour of the outer domain has been changed to transparent blue, to emphasize the other structural elements that undergo conformational changes upon CD4 binding. Carbohydrates are omitted for clarity. **b**, Same view of a homology model for the liganded SIV gp120 core, based on the HIV gp120 core structure in the CD4-bound conformation[14]. **c**, The unliganded gp120 core oriented so that the likely connection to gp41 faces 'downwards' as seen in the figure; gp41 is represented by the light-blue oval. The movements of the outer domain and of the V1–V2 stem upon CD4 binding are indicated by a blue and a red arrow, respectively. CD4 enters from 'above' (green arrow). **d**, The gp120 core in the CD4-bound conformation, oriented to correspond to the view in **c**, when the three-strand, inner-domain β-sheet (including β5, β7 and β25, in cyan) is aligned. This panel can also be viewed as the endpoint of the motions of the outer domain and V1–V2 stem indicated in **c**; the N and C termini still point downwards to gp41 (light blue), whereas the bridging sheet now faces upwards towards the target-cell membrane.

Surfaces and antigenicity

Mapping of epitopes that bind monoclonal antibodies has led to a distinction among 'neutralizing', 'non-neutralizing' and 'silent' faces of gp120 (refs 17, 25) (Fig. 5). These are thought to correspond, respectively, to epitopes on free gp120 that are exposed on the virion-associated gp120/gp41 trimer, epitopes on the free fragment that are buried on the trimer (probably at trimer interfaces or gp41 interfaces), and protein surfaces concealed by glycans. The extent of the differences between the unliganded and CD4-bound conformations now reinforces the further distinction between antibodies that recognize, stabilize, or induce one conformation and those that recognize the other. Appreciation of these two categories suggests the design of gp120 and gp160 locked in the unliganded state as selective immunogens.

Rearrangements during the transition between unliganded and CD4-bound conformations of gp120 have been invoked to explain the large negative entropy (and correspondingly large negative enthalpy) of CD4 binding[26]. Antibodies that recognize the receptor site on gp120 (so-called CD4bs antibodies) also have a large negative binding entropy and presumably require the same conformational transition[27]. This entropy cost reduces antibody and receptor affinity and provides one explanation for the failure of gp120 to induce potent neutralizing antibodies against conserved features of the CD4 and co-receptor sites[27]. A further explanation may be found in the unliganded conformation itself, which separates some of those conserved features by interspersing non-conserved elements (Fig. 4a) and which conceals others within a narrow cavity (Fig. 4b). Empirical relationships between thermodynamic parameters and conformational differences suggest that the measured entropy and enthalpy changes could either come from the ordering (upon CD4 binding) of about 100 residues or from the burial of about 10,000 Å^2 of molecular surface[26]. Surface-area estimates from our unliganded SIV gp120 core structure and that of CD4-bound HIV gp120 show, after compensating for glycosylation differences and for the presence of the 17b Fab fragment, that about 7,000 Å^2 of surface is sequestered from solvent when CD4

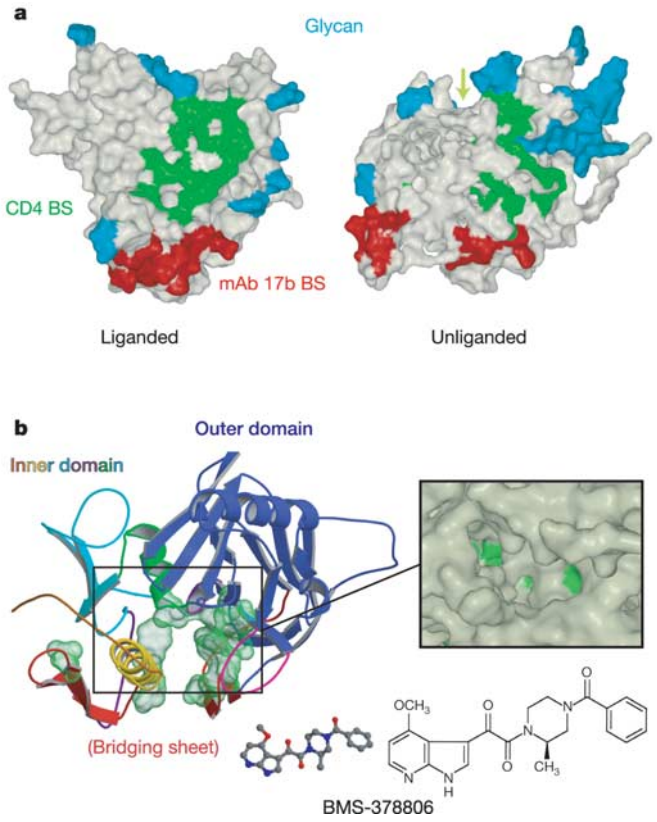

Figure 4 The binding sites (BS) of CD4 and monoclonal antibody (mAb) 17b (ref. 14), and the putative binding site of BMS-378806 (ref. 24), a small-molecule inhibitor of HIV-1 entry. **a**, Molecular surface representations of gp120 core structures in liganded and unliganded states. View as in Fig. 2a, b. Residues in direct contact with CD4 are in green; residues contacting human monocolonal antibody 17b, in red; carbohydrate, in light blue. Residues were chosen automatically from the HIV gp120 structure with the program HBPLUS[49], and were visualized with the programs LIGPLOT[50] and O. Contact criteria were quite generous, with residues defined as neighbouring if within 4.2 Å, and as hydrogen bonded if the hydrogen-to-acceptor distance was less than 3.0 Å and the donor-to-acceptor distance less than 3.6 Å. Green arrow indicates the mouth of the hydrophobic cavity shown in b. **b**, The unliganded gp120 core structure, as if viewed from the top in **a**. The residues lining a deep, hydrophobic cavity and corresponding to BMS-378806-selected resistance mutations in HIV gp120 are shown in surface rendering. In SIV gp120, these residues are Phe 94, Glu 95, Leu 107, Thr 272, Trp 391, Phe 398, Thr 439, Trp 440, Asn 446 and Val 485. The inset illustrates the entire cavity, with the visible resistance-mutation positions highlighted in green. The molecular formula of BMS-378806 is also shown, as well as a ball-and-stick representation on the same scale as the ribbon diagram.

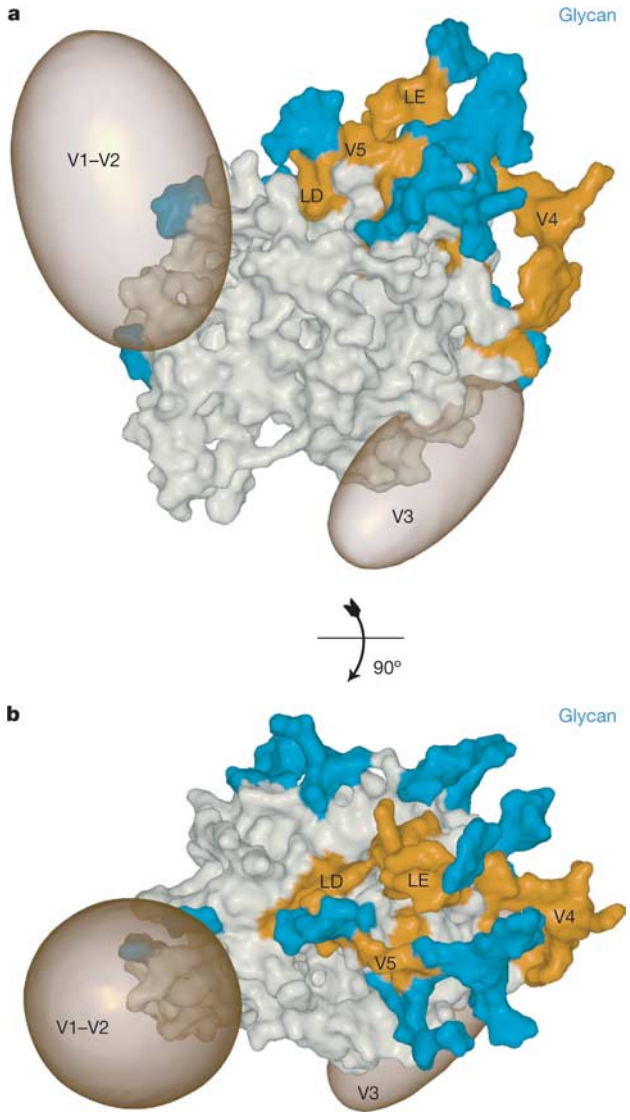

Figure 5 Surface structure of the glycosylated, unliganded gp120 core. **a**, Molecular surface of the unliganded gp120 core, viewed as in Fig. 4a. Carbohydrates are in light blue; surface loops LD and LE, and variable loops V4 and V5, in orange. Truncated variable loops V1, V2 and V3 are represented in transparent brown by mock surfaces derived from unrelated structures of comparable size. **b**, Top view of structure shown in **a**.

and gp120 associate, in reasonable agreement with the empirical prediction. There are likely also to be varying degrees of rigidification upon receptor binding, from a full disorder–order transition for residues 220–228 to reduced group motions for the inner-domain substructures.

The antibody b12 (ref. 28) is unusual in two ways: it has broad neutralizing activity and it has a markedly smaller negative entropy (as well as negative enthalpy) of binding than do other CD4bs monoclonals. Nonetheless, it probably recognizes the CD4-bound conformation, because the positions of alanine substitutions that alter its binding map to a contiguous surface on the CD4-bound gp120 structure, but map to a more distributed surface on the unliganded protein[29]. It is possible that it relies more heavily on outer-domain contacts than do other CD4bs antibodies and hence that it requires a less complete inner-domain transition.

There are examples of monoclonal antibodies with discontinuous epitopes that map jointly to V1–V2 and V3 (refs 30, 31). These two loops project away from each other in the structure described here, and it is not likely that a single antigen-combining region of an IgG could contact both simultaneously. Models for the Env trimer based on the unliganded conformation, described below, show that the V1–V2 loop from one subunit could approach or contact the V3 loop from another subunit. We suggest that proximity of these loops from two different subunits in the trimer, rather than proximity within a single gp120, may account for the joint epitope.

A drug-binding site

A recently discovered compound, BMS-378806, and some of its analogues, inhibit HIV-1 entry[24]. Resistance mutations have been used to identify the residues that are likely to contact these candidate drugs[32,33]. The sites of most of the resistance mutations lie in the deep, hydrophobic channel of unliganded gp120, as illustrated in Fig. 4b. They define a convincing binding site for molecules such as BMS-378806. The same residues do not line an obvious hydrophobic cavity in the CD4-bound structure. These observations suggest that the compounds inhibit entry by stabilizing the unliganded conformation. A group of fusion-inhibiting peptides[34] probably act in the same way. BMS-378806 does not inhibit SIV (ref. 24), as several of the HIV-1 resistance mutations are already present in the SIV gp120 sequence.

Conformation of gp120 in virion-associated trimers

We have created an approximate model for the unliganded trimer (Fig. 6). We have taken the structure described here as an approximation to each gp120 core on the unliganded trimer and applied the following criteria. First, the N and C termini of gp120 should point towards gp41, presumed to be centred on the three-fold axis, and roughly 'downwards' towards the viral membrane. Second, the N-linked glycans should be exposed, rather than buried at potential trimer contacts. Third, the outermost tips of the three gp120 subunits should project away from the three-fold axis, rather than cluster around it, to account for the appearance of stabilized SIV gp140 trimers in electron micrographs of negatively stained molecules[35]. We have applied these criteria qualitatively, by visual inspection, rather than attempting to optimize a quantitative formulation. A model that satisfies them is shown in Fig. 6.

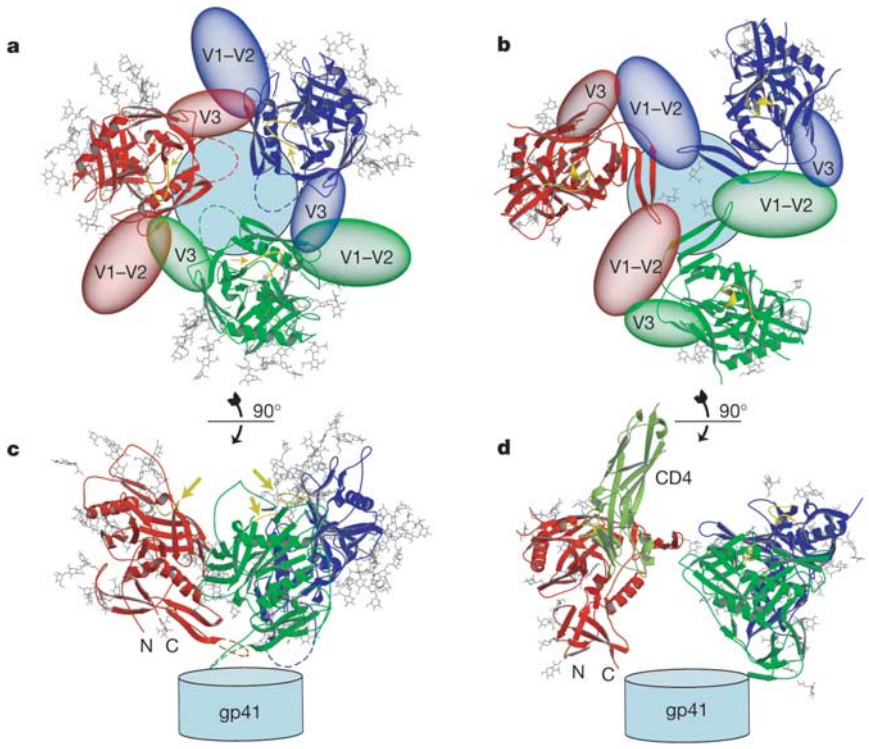

Figure 6 Proposed models for gp120/gp41 trimers in unliganded and CD4-bound conformations. **a**, A trimer in the unliganded conformation, viewed along the three-fold axis from outside the virion towards gp41. The polypeptide chain backbones are in ribbon representation; N-linked glycans are stick models; deleted V1–V2 and V3 segments are transparent balloons. The three monomers are in red, green and blue, respectively; the sugars, in grey. gp41 is shown as a circle in the rear. **b**, The same view of a gp120/gp41 trimer as in **a**, but in the CD4-bound conformation, generated by superposing the CD4-bound HIV gp120 structure onto the unliganded SIV gp120 subunits in panel **a**, assuming that the three-strand, inner-domain β-sheet remains roughly in place (see Fig. 3c, d). Structural elements depicted as in **a**; CD4 omitted for clarity. **c**, 'Side' view of the same model as in **a**; the red monomer is in an orientation similar to the one in Fig. 3c. The N and C termini of the gp120 core are labelled. gp41 is shown as a cylinder at the bottom. Green arrows indicate CD4-binding loops. **d**, Side view of the same model as in **b**, with the red monomer viewed as in Fig. 3d. The first two domains of CD4 are shown in light green on only one gp120 monomer. N and C termini of the gp120 core are labelled. gp41 is shown as a cylinder at the bottom.

Although this model is only one of a range of acceptable structures, its properties are probably representative of the whole set.

In this model for an unliganded trimer, the three receptor-binding sites face a bowl-like cavity around the three-fold axis at the 'top' of the molecule (opening away from the viral membrane). The CD4-binding loop is accessible, projecting over the tip of the β20–β21 ribbon. The V1–V2 stem is further recessed within the cavity, but its tip projects laterally outwards near the surface of a neighbouring gp120. A noteworthy feature of this model is the position of the disordered loop (residues from 220 to 228; dashed line in Fig. 6) connecting the V1–V2 stem to the inner-domain β-sheet. It lies near the three-fold axis, and we propose that in the unliganded trimer it contacts gp41. In the CD4-bound conformation, this loop is well ordered: the conformational change stretches it into an extended strand. Its displacement might contribute to weakening of the non-covalent gp120/gp41 association, known to accompany CD4 binding[36].

A further feature of our model is a potential interaction between V1–V2 of one subunit and V3 of another. In addition to accounting for the apparent joint epitope for certain antibodies, mentioned earlier, a role for both loops in stabilizing the unliganded trimer would account for a number of additional observations. TCLA virus and PBMC-adapted virus are sensitive to neutralization by soluble CD4. The main determinant of this phenotype seems to be V2 (ref. 37). Likewise, changes frequently found in CD4-independent strains are in the V1–V2 loop and within or close to HR1 of gp41 (ref. 38). These strains are also CD4-neutralization sensitive[39,40]. Both phenotypes could result from destabilization of the unliganded conformation and hence greater ease of access by soluble CD4 to its binding site. The same may be true of the three mutations reported in an early study of T-cell line adaptation, which lie near a domain contact within a single gp120, between the N terminus of α1 and a loop of the outer domain, but which are also disrupted by CD4 binding[41].

We have explained above that the most appropriate frame of reference for relating the CD4-bound conformation of the gp120 core to the unliganded conformation is one in which the inner-domain sheet, and the associated N and C termini, remain fixed. This assumption leads to the picture of a liganded trimer shown in Fig. 6b, d. The bridging sheet faces 'upwards', as required for interaction with the co-receptor. The orientation of the gp120 core with respect to the three-fold axis is similar to the one derived by Kwong *et al.*[42], using criteria related to the ones we outline above. One apparent difference, but within the range of uncertainty of the two approaches, is the orientation of CD4, which approaches nearly parallel to the axis of the trimer in our model, but at about 45° from the 'side' in the Kwong *et al.* model[42]. To account for electron microscopy images of negatively stained trimers[35], we have also displaced the gp120 core farther from the axis than did Kwong *et al.* We suggest that most of the three-fold contacts are made through gp41 and the N- and C-terminal segments of gp120, not present in either structure.

CD4 binding and commitment to the fusion-promoting conformational transition

Even in the absence of CD4, it is likely that gp120 will fluctuate towards the CD4-bound conformation; approach of CD4 could then stabilize this excursion and ultimately fix the liganded state. An initial contact might be made at the outer-domain CD4-binding loop, which is exposed in both conformations and which moves by only about 10 Å in the transition from one to the other, as modelled in Fig. 3. Our model predicts that other parts of gp120 will shift around the CD4-binding loop with much larger excursions (for example, the tip of the V1–V2 stem moves by over 40 Å). These displacements can occur as CD4 docks. In this process, the CD4-binding loop will curl around to present three residues of extended chain for antiparallel β-sheet hydrogen-bond interaction with the C″ strand at the edge of CD4 domain 1, as seen in the CD4-bound structure[14]. The last of these three residues, a conserved aspartic acid, also has a bidentate salt bridge with CD4 Arg 59. The critical CD4 residue Phe 43, at the beginning of the C″ strand, will then fit into a cavity between residues just distal to this aspartic acid and the edge of the bridging sheet. Reconfiguration of the loop joining the V1–V2 stem to the inner-domain sheet could, as suggested above, promote dissociation from gp41, liberating gp41 to commit to the fusion transition.

The CD4-bound structure determined by Kwong *et al.*[14] includes the 17b Fab, likely to be a surrogate for the co-receptor. Is there an intermediate structure, for the state with bound CD4, but no co-receptor or antibody? A possibility, suggested by the relatively independent shifts of individual secondary-structure elements between the unliganded and CD4-bound states, is that until co-receptor (or surrogate antibody) binds, the V1–V2 stem is not docked against the β20–β21 ribbon from the outer domain, and thus the bridging sheet is not properly formed. Repositioning of the V1–V2 stem, induced or fixed by the co-receptor, would then 'pull' on the loop that joins the stem to the inner-domain sheet, as outlined at the end of the preceding paragraph, and release constraints on gp41. □

Methods

Protein production and crystallization

SIV gp120 core protein was expressed in insect cells and purified by antibody affinity chromatography as described[21]. For large-scale protein production, 12 l of *Trichoplusia ni* (Hi-5) cells were infected with recombinant baculoviruses at a multiplicity of infection (MOI) of 2.5. The supernatant was harvested 72 h postinfection by centrifugation, concentrated, and loaded onto an antibody 17A11 affinity column. The protein was eluted and further purified by gel filtration chromatography on Superdex 200 (Pharmacia). Fractions containing SIV gp120 core were pooled and concentrated to an absorbance of 30 at 280 nm ($A_{280} = 30$). To produce selenomethionine-substituted gp120 core protein, infected Hi-5 cells were grown in medium supplemented with 75 mg l^{-1} selenomethionine (Sigma) and 5% FBS (Sigma), after starvation for 7 h in medium lacking methionine (JRH Biosciences). The protein was purified by the same procedure described above.

Crystals were grown in hanging drops from a mother liquor containing 15% PEG 6000, 100 mM sodium citrate, pH 5.0 and 8% PEG 400 at 20 °C. Crystals were briefly soaked in mother liquor before transfer to a solution containing 17% PEG 6000, 100 mM sodium citrate, 8% PEG 400 and 15% sucrose, in which they were flash-frozen in liquid nitrogen. Heavy atom derivatives were made by soaking crystals in mother liquor containing 1 mM of heavy atom compounds for 24 h at 20 °C.

Structure determination and trimer modelling

Crystal screening and data collection were performed at Cornell High Energy Synchrotron Source (CHESS) beamline F1 and Advanced Photon Source (APS) beamline 19ID. The best data (about 4 Å resolution) for native crystals and for crystals with bound K_2IrCl_6 or trimethyl lead acetate (TMLA) were collected at CHESS beamline F1. Data (to about 4.7 Å resolution) for SeMet crystals at peak wavelength were obtained at APS beamline 19ID. Denzo and Scalepack or HKL2000 were used for data processing[43]. One iridium site for the K_2IrCl_6 derivative was initially identified from isomorphous and anomalous differences using the program SOLVE[44]. Lead sites for the TMLA derivative and selenium sites for the selenomethionine-substituted protein were subsequently identified by difference Fourier analysis. The heavy atom sites from all the derivatives were refined together using the program SHARP[45].

Subsequent phase improvement by iterative cycles of phase combination, data sharpening, density modification, multi-crystal averaging, model building and heavily restrained refinement were performed using the CCP4 programs[46] and CNS[47]. Program O[48] was used for model building. N-linked glycans from high-resolution structures deposited in the Protein Data Bank (PDB) were compiled to generate a sugar library for studying sugar conformation. Some of the fucosylated N-linked glycans from the library were fit to experimental electron density maps. Heavily restrained refinement was performed using both CNS and Refmac[46]. Restraints include main-chain hydrogen-bond restraints, phi- and psi-restraints for helices, and harmonic restraints for all the glycans, main chain and side chains. Torsional simulated annealing and minimization refinement were performed in CNS. TLS refinement, followed by limited minimization with a flat bulk solvent model was implemented using the program Refmac 5. The final R_{free} and R_{work} are 38.8% and 38.5%, respectively. The relatively low data-to-parameter ratio prevented further improvement. We validated the structure by calculating averaged simulated-annealing omit maps[47], by comparing the final model with the experimentally phased map, and by calculating a model-based anomalous difference map for the SeMet sites. Full details of the structure determination and validation are given in ref. 21.

The trimer was modelled using a set of macros for O that allowed us to move one monomer freely; the two other monomers related by three-fold symmetry were generated and updated in a P3 cell. The model was judged by visual inspection based on the criteria described in main text. A trimer model for the CD-bound conformation was generated by aligning the three-strand inner-domain β-sheet as described in the caption to Fig. 6.

Received 13 October; accepted 22 December 2004; doi:10.1038/nature03327.

1. Wyatt, R. & Sodroski, J. The HIV-1 envelope glycoproteins: Fusogens, antigens, and immunogens. *Science* **280,** 1884–1888 (1998).
2. Allan, J. S. *et al.* Major glycoprotein antigens that induce antibodies in AIDS patients are encoded by HTLV-III. *Science* **228,** 1091–1094 (1985).
3. Veronese, F. D. *et al.* Characterization of gp41 as the transmembrane protein coded by the HTLV-III/LAV envelope gene. *Science* **229,** 1402–1405 (1985).
4. Center, R. J. *et al.* Oligomeric structure of the human immunodeficiency virus type 1 envelope protein on the virion surface. *J. Virol.* **76,** 7863–7867 (2002).
5. Dalgleish, A. G. *et al.* The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus. *Nature* **312,** 763–767 (1984).
6. Feng, Y., Broder, C. C., Kennedy, P. E. & Berger, E. A. HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor. *Science* **272,** 872–877 (1996).
7. Trkola, A. *et al.* CD4-dependent, antibody-sensitive interactions between HIV-1 and its co-receptor CCR-5. *Nature* **384,** 184–187 (1996).
8. Wu, L. *et al.* CD4-induced interaction of primary HIV-1 gp120 glycoproteins with the chemokine receptor CCR-5. *Nature* **384,** 179–183 (1996).
9. Sattentau, Q. J. & Moore, J. P. Conformational changes induced in the human immunodeficiency virus envelope glycoprotein by soluble CD4 binding. *J. Exp. Med.* **174,** 407–415 (1991).
10. Sattentau, Q. J., Moore, J. P., Vignaux, F., Traincard, F. & Poignard, P. Conformational changes induced in the envelope glycoproteins of the human and simian immunodeficiency viruses by soluble receptor binding. *J. Virol.* **67,** 7383–7393 (1993).
11. Rizzuto, C. D. *et al* A conserved HIV gp120 glycoprotein structure involved in chemokine receptor binding. *Science* **280,** 1949–1953 (1998).
12. Chan, D. C., Fass, D., Berger, J. M. & Kim, P. S. Core structure of gp41 from the HIV envelope glycoprotein. *Cell* **89,** 263–273 (1997).
13. Weissenhorn, W., Dessen, A., Harrison, S. C., Skehel, J. J. & Wiley, D. C. Atomic structure of the ectodomain from HIV-1 gp41. *Nature* **387,** 426–430 (1997).
14. Kwong, P. D. *et al.* Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature* **393,** 648–659 (1998).
15. Chan, D. C., Chutkowski, C. T. & Kim, P. S. Evidence that a prominent cavity in the coiled coil of HIV type 1 gp41 is an attractive drug target. *Proc. Natl Acad. Sci. USA* **95,** 15613–15617 (1998).
16. Rimsky, L. T., Shugars, D. C. & Matthews, T. J. Determinants of human immunodeficiency virus type 1 resistance to gp41-derived inhibitory peptides. *J. Virol.* **72,** 986–993 (1998).
17. Wyatt, R. *et al.* The antigenic structure of the HIV gp120 envelope glycoprotein. *Nature* **393,** 705–711 (1998).
18. Pollard, S. R., Meier, W., Chow, P., Rosa, J. J. & Wiley, D. C. CD4-binding regions of human immunodeficiency virus envelope glycoprotein gp120 defined by proteolytic digestion. *Proc. Natl Acad. Sci. USA* **88,** 11320–11324 (1991).
19. Wyatt, R. *et al.* Functional and immunologic characterization of human immunodeficiency virus type 1 envelope glycoproteins containing deletions of the major variable regions. *J. Virol.* **67,** 4557–4565 (1993).
20. Rud, E. W. *et al.* in *Vaccines 92: Modern Approaches to New Vaccines Including Prevention of AIDS* (eds Brown, F., Chanock, R. M., Ginsberg, H. S. & Lerner, R. A.) 229–235 (Cold Spring Harbor Laboratory, New York, 1992).
21. Chen, B., Vogan, E., Gong, H. Y., Wiley, D. C. & Harrison, S. C. Determining the structure of an unliganded and fully-glycosylated SIV gp120 envelope glycoprotein. *Structure (Camb.)* (in the press).
22. Kwong, P. D. *et al.* Probability analysis of variational crystallization and its application to gp120, the exterior envelope glycoprotein of type 1 human immunodeficiency virus (HIV-1). *J. Biol. Chem.* **274,** 4115–4123 (1999).
23. Thali, M. *et al.* Characterization of conserved human immunodeficiency virus type 1 gp120 neutralization epitopes exposed upon gp120–CD4 binding. *J. Virol.* **67,** 3978–3988 (1993).
24. Lin, P. F. *et al.* A small molecule HIV-1 inhibitor that targets the HIV-1 envelope and inhibits CD4 receptor binding. *Proc. Natl Acad. Sci. USA* **100,** 11013–11018 (2003).
25. Moore, J. P. & Sodroski, J. Antibody cross-competition analysis of the human immunodeficiency virus type 1 gp120 exterior envelope glycoprotein. *J. Virol.* **70,** 1863–1872 (1996).
26. Myszka, D. G. *et al.* Energetics of the HIV gp120–CD4 binding reaction. *Proc. Natl Acad. Sci. USA* **97,** 9026–9031 (2000).
27. Kwong, P. D. *et al.* HIV-1 evades antibody-mediated neutralization through conformational masking of receptor-binding sites. *Nature* **420,** 678–682 (2002).
28. Burton, D. R. *et al.* Efficient neutralization of primary isolates of HIV-1 by a recombinant human monoclonal antibody. *Science* **266,** 1024–1027 (1994).
29. Pantophlet, R. *et al.* Fine mapping of the interaction of neutralizing and nonneutralizing monoclonal antibodies with the CD4 binding site of human immunodeficiency virus type 1 gp120. *J. Virol.* **77,** 642–658 (2003).
30. Etemad-Moghadam, B. *et al.* Characterization of simian-human immunodeficiency virus envelope glycoprotein epitopes recognized by neutralizing antibodies from infected monkeys. *J. Virol.* **72,** 8437–8445 (1998).
31. Zwick, M. B. *et al.* A novel human antibody against human immunodeficiency virus type 1 gp120 is V1, V2, and V3 loop dependent and helps delimit the epitope of the broadly neutralizing antibody immunoglobulin G1 b12. *J. Virol.* **77,** 6965–6978 (2003).
32. Guo, Q. *et al.* Biochemical and genetic characterizations of a novel human immunodeficiency virus type 1 inhibitor that blocks gp120–CD4 interactions. *J. Virol.* **77,** 10528–10536 (2003).
33. Madani, N. *et al.* Localized changes in the gp120 envelope glycoprotein confer resistance to human immunodeficiency virus entry inhibitors BMS-806 and #155. *J. Virol.* **78,** 3742–3752 (2004).
34. Ferrer, M. & Harrison, S. C. Peptide ligands to human immunodeficiency virus type 1 gp120 identified from phage display libraries. *J. Virol.* **73,** 5795–5802 (1999).
35. Chen, B. *et al.* A chimeric protein of simian immunodeficiency virus envelope glycoprotein gp140 and *Escherichia coli* aspartate transcarbamoylase. *J. Virol.* **78,** 4508–4516 (2004).
36. Moore, J. P., McKeating, J. A., Weiss, R. A. & Sattentau, Q. J. Dissociation of gp120 from HIV-1 virions induced by soluble CD4. *Science* **250,** 1139–1142 (1990).
37. Pugach, P. *et al.* The prolonged culture of human immunodeficiency virus type 1 in primary lymphocytes increases its sensitivity to neutralization by soluble CD4. *Virology* **321,** 8–22 (2004).
38. Puffer, B. A., Altamura, L. A., Pierson, T. C. & Doms, R. W. Determinants within gp120 and gp41 contribute to CD4 independence of SIV Envs. *Virology* **327,** 16–25 (2004).
39. Kolchinsky, P., Kiprilov, E. & Sodroski, J. Increased neutralization sensitivity of CD4-independent human immunodeficiency virus variants. *J. Virol.* **75,** 2041–2050 (2001).
40. Puffer, B. A. *et al.* CD4 independence of simian immunodeficiency virus Envs is associated with macrophage tropism, neutralization sensitivity, and attenuated pathogenicity. *J. Virol.* **76,** 2595–2605 (2002).
41. Turner, S. *et al.* Resistance of primary isolates of human immunodeficiency virus type 1 to neutralization by soluble CD4 is not due to lower affinity with the viral envelope glycoprotein gp120. *Proc. Natl Acad. Sci. USA* **89,** 1335–1339 (1992).
42. Kwong, P. D., Wyatt, R., Sattentau, Q. J., Sodroski, J. & Hendrickson, W. A. Oligomeric modeling and electrostatic analysis of the gp120 envelope glycoprotein of human immunodeficiency virus. *J. Virol.* **74,** 1961–1972 (2000).
43. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276,** 307–326 (1997).
44. Terwilliger, T. C. & Berendzen, J. Automated MAD and MIR structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **55,** 849–861 (1999).
45. de la Fortelle, E. & Bricogne, G. Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.* **276,** 472–493 (1997).
46. Collaborative Computational Project, N. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50,** 760–763 (1994).
47. Brunger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54,** 905–921 (1998).
48. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47,** 110–119 (1991).
49. McDonald, I. K. & Thornton, J. M. Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.* **238,** 777–793 (1994).
50. Wallace, A. C., Laskowski, R. A. & Thornton, J. M. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng.* **8,** 127–134 (1995).

Supplementary Information accompanies the paper on **www.nature.com/nature**.

Acknowledgements We thank staff at CHESS beamline F1 and APS beamline 19ID for assistance, J. Hoxie of University of Pennsylvania, for hybridomas, and members of the Harrison/Wiley laboratory for discussion. The research was supported by a Scholar Award from the American Foundation for AIDS Research (to B.C.), by the NIH Innovation Grant Program for Approaches in HIV Vaccine Research (to S.C.H. and D.C.W.), and by an NIH HIVRAD grant (to Ellis Reinherz). S.C.H. is, and D.C.W. was, an Investigator of the Howard Hughes Medical Institute.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.C.H. (harrison@crystal.harvard.edu). Coordinates and structure factors have been deposited in the PDB, accession number 2BF1.