Structure, Vol. 13, 197–211, February, 2005, ©2005 Elsevier Ltd All rights reserved. DOI 10.1016/j.str.2004.12.004

Determining the Structure of an Unliganded and Fully Glycosylated SIV gp120 Envelope Glycoprotein

Bing Chen,[1] Erik M. Vogan,[1,2] Haiyun Gong,[1] John J. Skehel,[3] Don C. Wiley,[1,2,4] and Stephen C. Harrison[1,2,*]

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

Summary

HIV/SIV envelope glycoproteins mediate the first steps in viral infection. They are trimers of a membrane-anchored polypeptide chain, cleaved into two fragments known as gp120 and gp41. The structure of HIV gp120 bound with receptor (CD4) has been known for some time. We have now determined the structure of a fully glycosylated SIV gp120 envelope glycoprotein in an unliganded conformation by X-ray crystallography at 4.0 Å resolution. We describe here our experimental and computational approaches, which may be relevant to other resolution-limited crystallographic problems. Key issues were attention to details of beam geometry mandated by small, weakly diffracting crystals, and choice of strategies for phase improvement, starting with two isomorphous derivatives and including multicrystal averaging. We validated the structure by analyzing composite omit maps, averaged among three distinct crystal lattices, and by calculating model-based, SeMet anomalous difference maps. There are at least four ordered sugars on many of the thirteen oligosaccharides.

Introduction

The envelope glycoproteins of the primate immunodeficiency viruses (HIV and SIV) mediate attachment to a host cell and fusion of viral and host cell membranes. They project from the surface of the virion, and they induce both neutralizing and nonneutralizing antibodies (Burton et al., 2004; Wyatt et al., 1998). They are also the target of fusion inhibitors, exemplified by T-20/Enfuvirtide (Kilby and Eron, 2003; Wild et al., 1994), and of some more recently studied attachment inhibitors (Lin et al., 2003; Wang et al., 2003). The envelope glycoprotein is synthesized as a precursor (gp160), which oligomerizes and is subsequently cleaved into the receptor binding subunit, gp120, and the transmembrane and fusion subunit, gp41 (Allan et al., 1985; Veronese et al., 1985). Cleavage is necessary for activating the fusion-promoting capacity.

Sequences of gp120 from different immunodeficiency viruses show five relatively conserved regions (C1–C5) and five variable regions (V1–V5) (Starcich et al., 1986) (Figure 1A). The flexibility and variability of the protein surface, resulting from both heavy glycosylation and variable polypeptide-chain loops, are believed to have evolved to evade the host immune response (Wei et al., 2003). On mature virions, the gp120/gp41 complex is a noncovalently linked trimeric spike (Center et al., 2001, 2002). Gp120 binds sequentially to the viral receptor and coreceptor on the surface of host cells (CD4 and a chemokine receptor, respectively), and in response to these interactions undergoes conformational changes that initiate the fusion process and lead to its dissociation from the trimer (Dalgleish et al., 1984; Feng et al., 1996; Moore et al., 1990; Sattentau and Moore, 1991; Sattentau et al., 1993; Trkola et al., 1996; Wu et al., 1996). The N-terminal fusion peptide of gp41 is then available to insert into the target cell membrane, followed by further large conformational changes that bring the two membranes together (Chan et al., 1997; Weissenhorn et al., 1997, 1999).

The structure of the gp120 "core" (gp120 stripped of N- and C-terminal extensions, V1–V2, and V3: see Figure 1 for definition), bound with CD4 and the Fab of mAb 17b, has been determined for two different HIV-1 isolates (Kwong et al., 1998, 2000). The site on gp120 for mAb 17b includes residues that determine the coreceptor interaction, and the antibody footprint is thought to mark the coreceptor site (Rizzuto et al., 1998). Thus, the structure probably represents the released form of gp120. The gp120 core folds into two closely associated domains, termed "inner" and "outer" (Kwong et al., 1998). The base of the V1–V2 variable region (the "V1–V2 stem"), in the inner domain, and a β-hairpin that projects from the outer domain form a four-strand "bridging sheet," which fixes the relative orientations of the two domains (Kwong et al., 1998). This sheet contributes to binding of both CD4 and mAb 17b, and it probably contains many or most of the residues critical for coreceptor interaction (Rizzuto and Sodroski, 2000; Rizzuto et al., 1998).

Carbohydrates of glycoproteins are a major source of surface flexibility, and glycoproteins with a large number of glycans are difficult to crystallize in their fully glycosylated form. We have determined the structure of a fully glycosylated SIV gp120 envelope glycoprotein in an unliganded conformation by X-ray crystallography at 4.0 Å resolution. Crystals of SIV gp120 diffract quite weakly, presumably due to heavy glycosylation and small crystal size. The low-resolution of our data and the lack of noncrystallographic symmetry posed considerable challenges for the structure determination. We describe here solutions to some of the experimental and computational difficulties, as our experience in de-

*Correspondence: harrison@crystal.harvard.edu
[4]Deceased

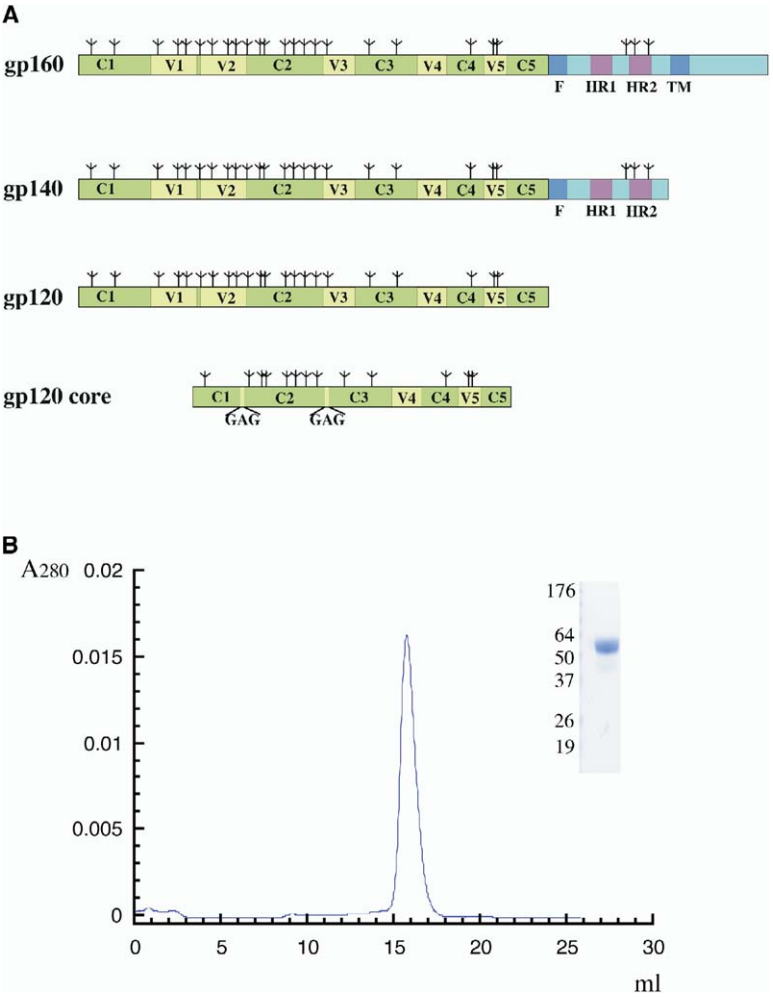

Figure 1. SIV Envelope Glycoproteins and Production of SIV gp120 Core Protein

(A) Schematic representations of SIV envelope glycoproteins: gp160, the full-length precursor; gp140, the uncleaved ectodomain of gp160; gp120, the surface subunit; and gp120 core, the truncated protein used for crystallographic studies. Glycans are represented by tree-like symbols. Various segments of gp120 and gp41 are designated as follows: C1–C5, conserved regions 1–5; V1–V5, variable regions 1–5; F, fusion peptide; HR1, heptad repeats 1; HR2, heptad repeats 2; TM, transmembrane segment.

(B) SIV gp120 core protein was purified by mAb 17A11 affinity chromatography from supernatants of insect cell culture, and then resolved by gel-filtration chromatography using a Superdex 200 column. Molecular weights were calculated based on a standard curve plotted using known proteins. Peak fractions were pooled and analyzed by SDS-PAGE (inset).

termining the structure of this heavily glycosylated protein may be useful for work on other crystallographic problems with limited resolution and poor initial phase information. Recording data adequate for structure analysis required optimal conditions at a synchrotron source. Initial experimental phases were gradually improved by iterative phase combination, data sharpening, density modification, multicrystal averaging, and model building. Most of the thirteen oligosaccharides have at least four well-ordered sugars. They cluster into networks, in some cases extending to decorate the protein surface of crystallographically related molecules.

Results

Expression and Purification of SIV gp120 Core Protein from Insect Cells

The surface flexibility of HIV/SIV envelope glycoproteins, due both to heavy glycosylation and surface-exposed variable loops (V1–V5; Figure 1A), has been a principal obstacle to crystallographic studies (Kwong et al., 1999). To produce an SIV gp120 protein suitable for X-ray crystallographic analysis, we have deleted 43 and 22 residues from the N- and C- termini of SIV gp120, respectively, following work on the HIV-1 gp120 core (Kwong et al., 1999; Wyatt et al., 1993), and we have substituted short linkers (GAG) for the V1-V2 and V3 loops (Figure 1A). Two extra residues (His-Met) were introduced by the restriction site (Nde I) at the N terminus. We designate this protein, which still retains 13 glycosylation sites, the SIV gp120 core protein. It was produced in insect cells using a baculovirus expression system and purified from cell supernatants by immunoaffinity chromatography using monoclonal antibody 17A11 (Edinger et al., 2000), followed by gel filtration. SIV gp120 core protein elutes from a Superdex 200 column as a sharp peak, which corresponds to a 60 kDa globular protein (Figure 1B). Analysis by SDS-PAGE showed that it migrates as a single, somewhat diffuse protein band, also with an apparent molecular mass of 60 kDa. The calculated mass of the polypeptide chain is about 39 kDa. The larger apparent size of the protein is due to heavy glycosylation, as indicated by the reduced size of the protein when treated with PNGase F, Endo F, or Endo H (data not shown).

We also prepared selenomethionine-substituted SIV gp120 core protein in order to obtain additional experi-

mental phase information, following the protocol described in Experimental Procedures.

Crystallization of the Fully Glycosylated SIV gp120 Core Protein

Initial attempts to crystallize SIV gp120 core protein were carried out using simple grid screens (concentration of precipitants versus pH) with polyethylene glycol 4000 or ammonium sulfate as precipitants. Some first hints of crystals appeared in hanging drops in which the protein had been mixed, at 1:1 ratio, with a mother liquor containing 15% polyethylene glycol 4000 and 100 mM sodium citrate, (pH 5). Fine-tuning of the crystallization conditions produced roundish crystals (Figure 2A), which grew to about 150 microns in the largest dimension in 2–3 weeks, but which yielded only a few reflections at very low (15 Å) resolution. We used additive screens (Hampton Research) to identify polyethylene glycol (PEG) 400 as useful, and we subsequently obtained tetragonal crystals by including 8% PEG 400 in the crystallization solution (Figure 2B). These crystals gave measurable diffraction to 4.0 Å at a synchrotron source (see below). Analysis of washed crystals by SDS-PAGE confirmed that they contained SIV gp120 core protein with no apparent degradation during the crystallization process (Figure 2E, lane 9).

Further attempts to improve crystalline order included varying the size of the polyethylene glycol precipitant (PEG 6000 gave slightly larger crystals than smaller or larger PEGs), the buffer, and other parameters. When 100 mM sodium acetate, (pH 5), was used instead of sodium citrate, elongated crystals appeared (Figure 2C), but these apparently trigonal crystals gave diffraction to spacings no better than 15 Å, even at the brightest synchrotron beamlines. Deglycosylation with PNGase F or Endo F prevented crystallization under all conditions examined. Use of Endo H, which yields only partial deglycosylation, produced gp120 core protein that crystallized from 2 M ammonium sulfate, (pH 6) (Figure 2D), but the crystals diffracted only to spacings of about 10 Å even when using crystals that grew to 100 microns in the longest dimension. In the case of HIV gp120 core protein, deglycosylation was essential for obtaining satisfactory crystals (Kwong et al., 1999), but this observation is evidently not a general rule. We also attempted to crystallize a number of complexes with Fab fragments, derived from mAbs 2C3, 2C9, 17A11, 3E9, 4E11, 11F2, 7D3, 8C7, and 5B11 (Edinger et al., 2000; Kim et al., 2001). None gave crystals suitable for diffraction experiments.

Diffraction Measurements

The SIV gp120 core protein crystals diffract very weakly, even allowing for their relatively small size. Of the several crystal forms we have obtained, the tetragonal crystals (Figure 2B) have yielded the best diffraction data. We could detect no diffraction from these crystals using a laboratory X-ray source, however, and long exposure times were required at synchrotron sources. All diffraction data were collected at CHESS beamline F1 and APS beamline 19ID. Optimization of crystal freezing conditions and beamline setup at the synchrotron allowed us to observe reflections to spacings of about 3.8 Å (Figure 3). We found 15% sucrose to be the best cryoprotectant. Dehydration by gradually increasing the PEG concentration did not improve the order, nor did annealing in the cryostream. A small beamstop (about 1 mm diameter) placed very close the crystal (about 1 cm) and a large crystal-to-detector distance (400 mm) were both essential for reducing background sufficiently to detect reflections beyond 4.5 Å. One of our best crystals gave weak diffraction with few overloads even after 3 min of exposure at CHESS F1 (Figure 3A), but the high-resolution reflections gradually faded into diffuse scattering in the water ring, suggesting that the crystals may be ordered to beyond 3.8 Å. Use of smaller nylon loops for freezing, to reduce the amount of solvent surrounding the crystal, did not decrease the water-ring intensity sufficiently to allow us to push the resolution beyond it. Our best data set for native crystals (for which we chose a 4.0 Å resolution cutoff based on a S/N ratio of 1.3, Bc44n in Table 1) was collected at CHESS F1. Data collected at APS 19ID were comparable to those from CHESS, despite the much brighter beam, because it was not possible to take further steps to reduce the relative strength of the water ring. The programs Denzo and Scalepack were used for integrating and scaling the data (Otwinowski and Minor, 1997). The crystals belong to space group $P4_32_12$ (a = b = 108.1 Å, c = 117.7 Å), with one molecule per crystallographic asymmetric unit and 60% solvent content.

Initial Phase Determination and Phase Improvement

Attempts to determine the structure by molecular replacement using various search models derived from the HIV gp120 core structure were unsuccessful. We screened (at CHESS F1) hundreds of crystals soaked with over 60 heavy atom compounds for isomorphous heavy-atom derivatives. Crystals soaked with either K_2IrCl_6 (Ibc1 and Sbc10L in Table1; Cbc4 in Figure 3B) or trimethyl lead acetate (TMLA) (Ibc44 in Table1) were isomorphous to unsoaked crystals and produced data sets from which heavy atom binding sites could be determined. One Ir site could be identified independently from isomorphous differences and anomalous differences (the wavelength at CHESS F1 was 0.916 Å, nearly optimal for single-wavelength anomalous dispersion from the L_I edge of Ir), using the program SOLVE (Terwilliger and Berendzen, 1999). One Pb site was found by difference Fourier analysis using phases calculated from the K_2IrCl_6 derivative (Cbc4). The Ir and Pb sites were crossvalidated using difference Fouriers. A third derivative could be produced by soaking crystals with $IrCl_3$ (Hbc7 in Table 1); those crystals suffered from significant nonisomorphism (cell constants a = b = 107.1 Å and c = 119.6 Å), and their contribution to the MIR phases was small due to the Ir site shared with the K_2IrCl_6 derivative (Ibc1 or Cbc4). When crystals were soaked with cadmium iodide, cell constants changed significantly (a = b = 110.9 Å and c = 111.9 Å); data collected from these crystals were quite useful for multicrystal averaging at a later stage (Hbc2/3 in Table 1; see below). We also obtained small SeMet-substituted crystals (about 70 microns in the largest dimension). The best data set (to about 4.7 Å resolution; Sb38 in Table 1) for SeMet crystals at the peak wavelength

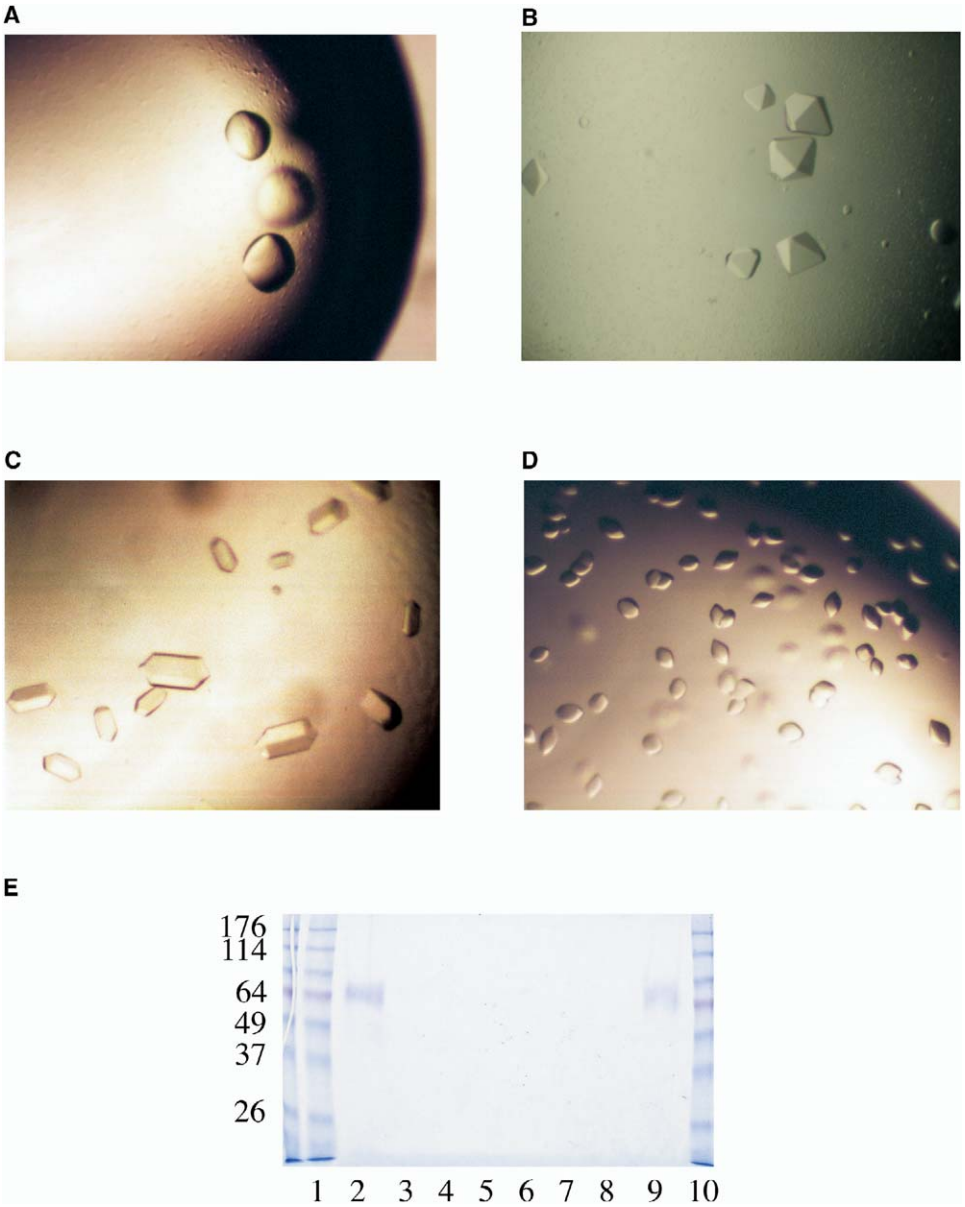

Figure 2. Crystallization of a Fully Glycosylated SIV gp120 Core Protein

The fully glycosylated SIV gp120 core protein crystallized in hanging or sitting drops under various conditions, with mother liquor containing 15% PEG 4000, 100 mM sodium citrate, pH 5.5 at 20°C (A); 15% PEG 6000, 100 mM sodium citrate, (pH 5.0), 8% PEG 400 at 20°C (B); 15% PEG 6000, 100 mM sodium acetate, (pH 5.0), 8% PEG 400 at 20°C (C); the protein deglycosylated by Endo H with mother liquor containing 2 M ammonium sulfate, 100 mM MES, pH 6.0 at 18°C (D).

(E) Three crystals of about 100 microns from (B) were carefully washed 6 times with mother liquor and then dissolved in SDS-loading buffer. The sample (lane 9) was then analyzed along with the drop (lane 2) and 6 washes (lanes 3 to 8) by SDS-PAGE.

(0.97935 Å) was obtained at APS 19ID. Eight selenium sites were subsequently identified by anomalous difference Fourier analysis using phases combined from the K_2IrCl_6, $IrCl_3$, and TMLA derivatives.

We refined the heavy-atom sites from all four derivatives together using the program SHARP (de la Fortelle and Bricogne, 1997). The resulting electron density map after density modification (using SHARP) allowed us to fit a partial model of the outer domain from the HIV gp120 core (Kwong et al., 1998). The long α-helix and a surrounding β sheet of the outer domain matched nicely with the map, as shown in Figure 4A. In addition, this fit placed each of Met 308, Met 356, Met 390, Met 402, and Met 481 in the vicinity of one selenium site each. The three sites that correspond to Met 390, Met 402, and Met 481 were too close to each other to be assigned definitively at the time, but the agreement clearly validated our initial experimental phases. Strong

A

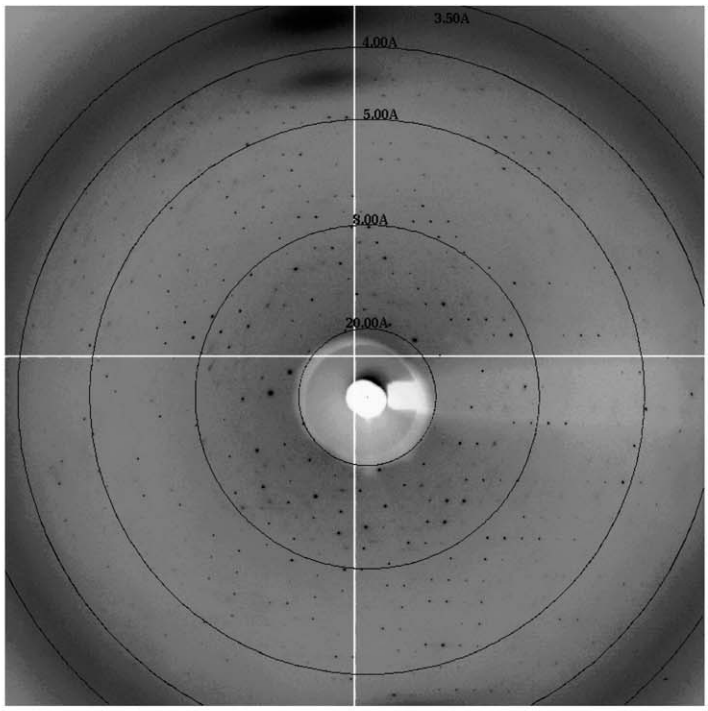

Figure 3. Diffraction from SIV gp120 Core Protein Crystals

(A) A diffraction image was recorded at CHESS beamline F1 using an ADSC Quantum-4 CCD detector with one of our best tetragonal crystals soaked with 1 mM K_2IrCl_6. A beam stop of 1 mm diameter, a large crystal-to-detector distance (at least 400 cm) and a long exposure time (3 min) were required to increase the signal-to-noise ratio. Reflections at 3.8 Å are clearly visible, but reflections at higher resolution fade into the diffuse scatter of the water ring.

(B) Inset table, crystallographic statistics of the complete data set from the crystal discussed in (A).

B

Data set	Cbc4
Compound	K_2IrCl_6
X-ray source	CHESS F1
Wavelength (Å)	0.916
Resolution limits (Å)	23.9 to 3.9
Cell constants (Å/°)	a=b=108.50, c=117.94
	α=β=γ=90.0
Mosaicity (°)	0.45
Measured reflections	85267
Unique reflections	5916
Mean I/σ(I)	10.2 (2.5)
R_{symm} (%)	10.9 (52.7)
Completeness (%)	85.9 (88.0)
Redundancy	7.3

features branching out from the polypeptide-chain density at several glycosylation sites provided further evidence for our initial fit.

Model Building and Refinement

Even from the electron density map calculated with experimental phases, we could see that the inner domain of unliganded SIV gp120 differs greatly from that of CD4 bound HIV gp120. We therefore focused our initial efforts on building the outer domain as accurately as possible, guided by the HIV model, in order to improve the phases. We took a polyalanine model of the portion of the HIV outer domain that fit the density well and used it as a starting point. Experimental data fall off sharply with increasing resolution (Figure 4B), and side chain density was apparent only after data were sharpened by applying a B factor of –200 to the amplitudes. Sequence registration was determined by alignment with the HIV model, selenium sites, and glycosylation sites. We used iterative cycles of phase combination (partial model phases and experimental phases), density modification, data sharpening, model building, and limited model refinement to improve and extend the initial model. After we had built about 70% of the structure, including some of the inner domain, the model was placed, by both rigid body refinement and molecular replacement, into the three crystal lattices available for multicrystal averaging: native (Bc44n), IrCl3-soaked

Table 1. Data Collection, Phasing, and Refinement Statistics

Data Set	Bc44n	Ibc1	Sbc10L	Sb38	Ibc44	Hbc7	Hbc2/3
Compound	Native	K_2IrCl_6	K_2IrCl_6	SeMet	TMLA	$IrCl_3$	CdI_2
Resolution limits (Å)	25.6 to 4.00	18.6 to 4.25	48.8 to 7.00	39.8 to 4.70	18.6 to 4.77	16.7 to 4.29	18.5 to 4.22
Cell constants (Å/°)	a = b = 108.05, c = 117.70	a = b = 107.43, c = 118.14	a = b = 106.90, c = 117.84	a = b = 107.49, c = 118.21	a = b = 109.40, c = 117.74	a = b = 106.95, c = 119.55	a = b = 110.86, c = 111.87
	$\alpha = \beta = \gamma = 90.0$	$\alpha = \beta = \gamma = 90.0$	$\alpha = \beta = \gamma = 90.0$	$\alpha = \beta = \gamma = 90.0$	$\alpha = \beta = \gamma = 90.0$	$\alpha = \beta = \gamma = 90.0$	$\alpha = \beta = \gamma = 90.0$
Mosaicity (°)	0.35	0.57	0.58	0.58	0.31	0.47	0.79
Measured reflections	100402	186459	20576	108316	73691	56384	39052
Unique reflections	6216	5132	1076	3859	3699	4367	4548
Mean I/σ(I)	12.0 (1.3)	15.4 (5.5)	15.5 (5.9)	13.3 (1.9)	10.9 (4.3)	9.6 (3.4)	5.7 (1.7)
R_{symm} (%)	8.3 (81.7)	11.8 (43.5)	9.0 (28.4)	10.6 (85.5)	10.0 (46.5)	12.1 (45.0)	13.3 (40.9)
Completeness (%)	98.0 (94.3)	97.4 (100.0)	87.3 (88.4)	98.2 (100.0)	95.2 (96.8)	85.3 (90.0)	84.1 (72.6)
Redundancy	4.5	13.8	7.1	6.8	6.7	6.4	2.9
Phasing							
Sites		2	2	7	2	1	
R_{merge}		0.169	0.205	0.110	0.177	0.463	
R_{Cullis} (iso/ano)		0.746/0.926	0.878/0.874	1.080/0.956	0.888/1.000	0.894/0.916	
Phasing power (iso/ano)		0.763/0.618	0.617/0.966	0.296/0.357	0.161/0.109	0.418/0.595	
FOM	0.25/0.32						
Refinement							
Residues built	304 (316)	B-Factor values		Ramachandran plot statistics			
Sugars built	50						
R_{free} reflections	300 (5%)	<model B>	128.8	Most favoured regions		62.1% (167)	
FOM	0.642	Patterson B	101.2	Generously allowed regions		5.9% (16)	
Bond lenghts (Å) /angles (°) rmsd	0.014/2.067			Disallowed regions		2.2% (6)	

All data were collected at λ=0.916 Å at beamline CHESS F1, except data set SB38 which was collected at λ = 0.97935 Å at beamline APS 19ID.

$R_{symm} = \Sigma(|I-\langle I\rangle|) / \Sigma(I)$ where I is the observed reflection intensity.

$R_{merge} = \Sigma(|F_{ph} - F_p|) / \Sigma(F_p)$ where F_p and F_{ph} are the observed native and derivative structure factors, respectively.

$R_{Cullis} = \Sigma(||F_{ph}|\pm|F_p||) - |F_h|| / \Sigma||F_{ph}| - |F_p||$ where F_p and F_{ph} are the observed native and derivative structure factors, respectively, and F_h is the (calculated)heavy-atom structure factor.

Phasing power = F_h / E where F_h is the (calculated) heavy-atom structure factor and E is the phase-integrated lack-of-closure.

$R_{cryst} = \Sigma(||F_{obs}| - |F_{calc}||) / \Sigma(|F_{obs}|)$ where F_{obs} (which equals F_p) is the observed native structure factor and F_{calc} is the structure factor calculated from the model.

(Hbc7), and CdI-soaked (Hbc2/3) crystals. Both rigid-body refinement and molecular replacement gave the same results, indicating that there was little global shift of the model in any of the lattices. We determined transformation matrices from the models placed in each of the different crystal lattices and carried out multicrystal density averaging, using a C-shell script to run the CCP4 programs MAPROT for averaging densities, RSTATS for scaling calculated and observed structure-factors, and FFT for map calculation (Collaborative Computational Project, 1994; Stein et al., 1994). This procedure allowed us to monitor every step of the computation. Six cycles of map averaging were usually enough to ensure phase convergence. When phases from a partial model containing only the outer domain were used, the Fo map showed discontinuous density for the long α helix in the inner domain (Figure 5A in green). With the same set of phases after multicrystal averaging, the density for the α helix became more continuous and even showed some features for side chains, indicating useful phase improvement (Figure 5A in purple).

With many cycles of averaging, model rebuilding, and gradually expanding the averaging mask when necessary, additional features emerged. The V4 variable loop, probably quite flexible in solution, makes crystal contacts with its counterpart from a symmetry-related molecule and appears ordered in the crystal. The density for most glycans was tube-like before multicrystal averaging, but in averaged maps a blob of density branched from the innermost N-acetyl glucosamine, consistent with reports that N-linked glycans on glycoproteins expressed in *T ni.* (Hi-5) cells are usually (α1-6) fucosylated at the protein-proximal N-acetyl glucosamine (Hsu et al., 1997). We assembled a library of fucosylated N-linked glycans from high-resolution structures deposited in the PDB, for fitting into the electron density maps (Figure 5B). We found density for sugars at all 13 glycosylation sites. A few glycans have ordered saccharides even beyond the first branching site because of crystal contacts, while for others only the first one or two sugars are visible.

In this manner, we located most of the inner domain secondary-structure elements, including two α helices, a β sheet, and four glycans, as well as the four strands of the bridging sheet, these last separated into two β-ribbons. Three potential selenium sites not in the outer domain had originally been identified from peaks in the anomalous difference map. Two of these sites, separated by just 6 Å, belong to the adjacent residues Met 255 and Met 256, in a one-turn α-helical configuration that connects inner and outer domains. Our current model does not account for the third SeMet anomalous difference peak, which could be a result of time-dependent radiation decay (Ravelli and McSweeney, 2000) or simply a consequence of poor isomorphous phases.

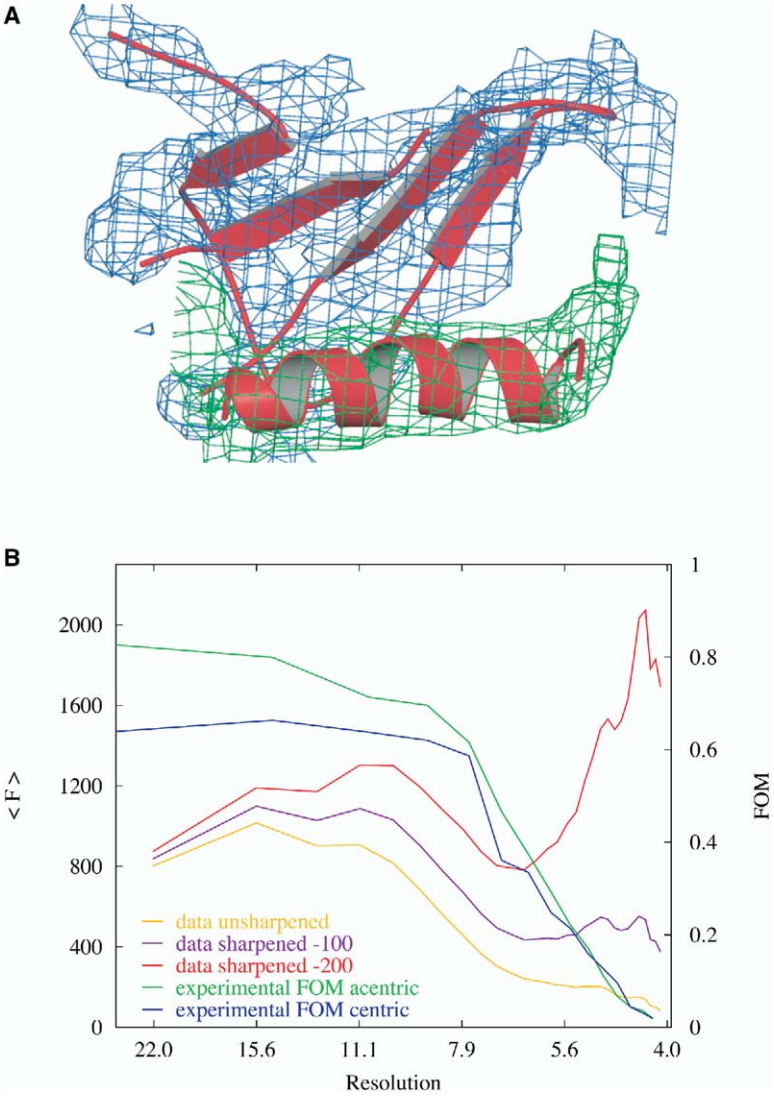

Figure 4. Fit of Part of the Outer Domain of HIV gp120 to the Electron Density Map from Experimental Phases

(A) Partial model of the outer domain from HIV gp120 core structure fit to electron density map computed from experimental phases after density modification. The long α-helix and a surrounding β sheet of the outer domain are shown as a ribbon diagram in red, density for the helix in green, and density for the sheet in blue.

(B) Plots of unsharpened and sharpened amplitudes and experimental figures of merit (FOM) versus resolution. Experimental amplitudes and FOM fall off sharply with increasing resolution, effectively reducing the resolution of the electron density map in (A) to 6.0 Å, despite inclusion of data to 4.0 Å resolution in the map calculation. Data sharpening by a B factor of –200 was necessary to reveal side chain density features.

Model-based phases later yielded anomalous difference peaks corresponding to the remaining two SeMet positions (see below). The V1-V2 stem (two of the strands that form the bridging sheet) appeared disordered until an averaged map was calculated at lower resolutions (5.0, 5.5, or 6.0 Å). When the contour level of these maps was set slightly lower than 1σ, density in the most likely region for the V1-V2 stem became apparent. The density matched the V1-V2 stem from the HIV gp120 core model; the stem is clearly quite rigid as a result of a disulfide bond at either end (Kwong et al., 1998).

The low data-to-parameter ratio (the ratio of unique X-ray reflections to the number of atoms in the model is less than 2) influenced efforts to refine the model. We carried out heavily restrained refinement calculations, using both CNS and Refmac 5 (Brunger et al., 1998; Murshudov, 1997). Main chain hydrogen-bond restraints and phi-, psi-restraints for helices were imposed in all CNS refinement runs, in order to prevent stereochemical distortion. As a result, R_{free} and R_{cryst} remained very similar. Harmonic restraints were also imposed for the protein main chain and side chains and for the glycans, varying the parameters so that outer mannoses were held less tightly than inner N-acetyl glucosamines. During model building, we used CNS to improve model phases. After the model was complete, two rounds of rebuilding based on the updated averaged map were carried out without any refinement, giving a model with R_{free} close to 50%. Torsional simulated annealing and minimization refinement in CNS, using native data from 12.0 to 4.0 Å without bulk solvent correction, brought the R_{free} to 43%.

B-factor refinement in CNS failed, probably from algorithmic instability due to insufficient experimental data. (Note that B-factor refinement at 4.7 Å resolution for p97 could succeed, because of noncrystallographic symmetry; DeLaBarre and Brunger, 2003). We therefore used TLS refinement as implemented in Refmac 5 to derive a better temperature factor model. A single TLS group was defined for the whole model, but analysis with up to six groups (five inner-domain "domainlets"

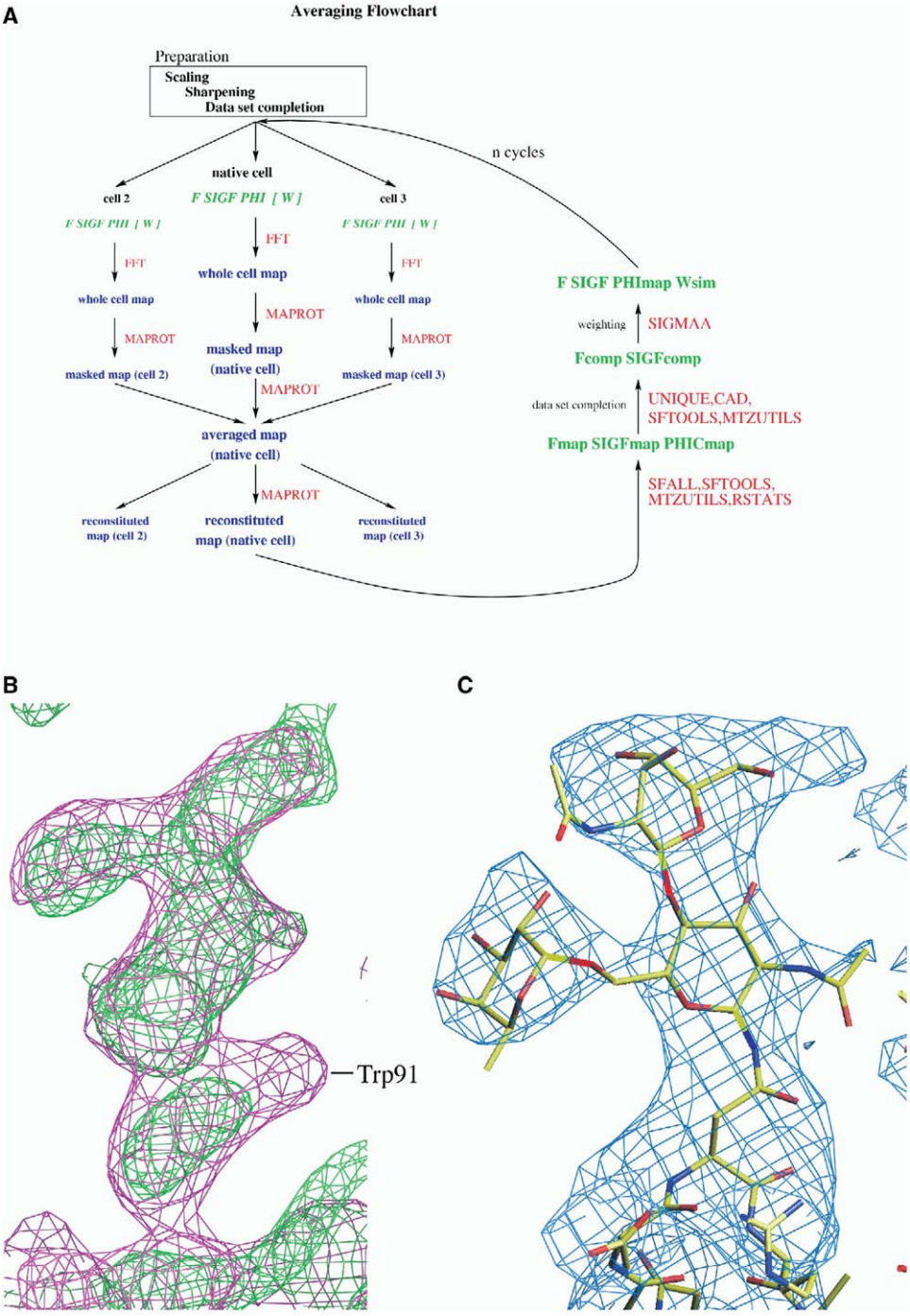

Figure 5. Phase Improvement by Multicrystal Averaging

(A) Flow chat for multicrystal averaging using CCP4 programs (Collaborative Computational Project, 1994).

(B) Phases from a model containing only part of the outer domain and experimental amplitudes were used to calculate an electron density map with a mask including both inner and outer domains. The density for the long α helix in the inner domain is shown before multicrystal averaging in green and after multicrystal averaging in purple. The density for Trp91 after averaging is as indicated. Both maps are contoured at 1σ.

(C) Density after multicrystal averaging is shown for glycan 294. In the averaged maps, a blob of density buds out of the innermost N-acetyl glucosamine residue, revealing the position of an α(1-6) fucosyl group.

plus the outer domain) gave essentially the same results. TLS refinement, followed by limited minimization with a flat bulk solvent model, yielded a model with R_{free} of 38.8% using native data from 26.0 to 4.0 Å. The solvent model was optimized by changing probe and shrink radii (the best values, 3.0 Å and 2.8 Å, respec-

tively, agree with those chosen during CNS refinement), and the geometry was strongly restrained. Some details of the TLS analysis are provided in Table 1. We further validated the TLS results by carrying out B-domain refinement in CNS and in Refmac 5, using each of the "domainlets" from the inner domain plus the outer domain as grouped structures (Supplemental Figure S1 available with this article online). The V1-V2 stem is substantially more mobile than the other structural elements.

Model Validation

Heavy Atoms

Isomorphous difference maps with model-based phases accurately return the positions of irridium and lead atoms in the derivatives (Supplemental Table S1). The TMLA sites are in hydrophobic pockets in the inner domain. The principal $IrCl_6^{2-}$ site (Ir1 in Supplemental Table S1) is adjacent to the second N-acetyl glucosamine of the glycan attached to Asn 459, in a pocket bounded both by sugars and by aromatic side chains (Trp 344 of the polypeptide chain to which the glycan belongs and Trp 342 and Tyr 428 of a crystallographically related chain). There are nearby positive charges that might neutralize the negative charge on the $IrCl_6^{2-}$ ion. The secondary $IrCl_6^{2-}$ site is likewise in a cavity amidst N-acetyl glucosamines, with a nearby arginine side chain. Thus, the heavy-atom sites, including the TMLA positions in the inner domain, have the chemical characteristics expected for the ligand they accommodate.

Selenium Positions

We calculated an anomalous difference map with model-based phases, using only those anomalous differences for which both amplitudes exceeded 2σ (Supplemental Figure S2). Of the nine selenium sites (including SeMet 65, introduced by the cloning strategy), seven—SeMets 65 and 251 in the inner domain, SeMets 265 and 266 in the connection between inner and outer domains, and SeMets 308, 390, and 403 in the outer domain—have strong corresponding peaks (Supplemental Table S1). SeMets 356 and 481, opposite each other in the hydrophobic core of the outer domain, have weaker peaks, but they are clearly seen in residual maps calculated during initial determination of Se locations, and they refine with reasonable occupancies (Supplemental Table S1). Five of the nine SeMet sites, including SeMet 65 and SeMet 251 that mark inner-domain elements, are included among the top twelve peaks. Two of the remaining three peaks, SeMets 265 and 266 (unambiguously identified in the initial anomalous difference map, Supplemental Table S1, and robustly refined in SHARP), may tend to cancel each other at the approximately 5 Å resolution cutoff of the difference map. Clarity of the model-based anomalous difference map is probably limited by phase differences between the slightly nonisomorphous SeMet and native lattices (see Table 1) and by modest accuracy of the SeMet diffraction data.

Comparison with Initial Experimental Map

Having built the inner domain, we looked back to compare it with a map calculated using phases derived directly from isomorphous and anomalous differences and hence completely unbiased by the model (Supplemental Figure S3, column A). The long helix, $\alpha1$, the sheet ($\beta5$, $\beta7$, $\beta25$), and the short helix ($\alpha5$) of the inner domain fit very well into the experimental density. The position of SeMet 251, in the $\beta5$–$\beta7$ loop, further anchors the sheet. One of the two bridging-sheet ribbons ($\beta20$–$\beta21$) has somewhat broken density. Only the other bridging-sheet ribbon ($\beta2$–$\beta3$, the V1–V2 stem) is not well represented in the experimental map.

Omit Maps

We calculated composite simulated-annealing omit maps using CNS. Figure 6 shows that all the secondary-structure elements in the inner domain are well defined, further confirming our general interpretation. Certain regions, mostly in loops, have weak density or breaks in the composite map. For example, density is discontinuous at the link in the inner domain between the N terminus and $\alpha1$, each of which lies in strong density (Figure 6F). The N terminus is at a crystal contact, and its position (with respect to the rest of the protein) appears to vary among the crystal forms we used for multicrystal averaging. Density for the glycan at Asn71 supports our interpretation of the sequence register for the N-terminal segment, as does the peak for SeMet65, but at the current resolution, simulated annealing composite omit maps are not powerful enough to guide further rebuilding.

Overall Assessment of the Inner Domain Structure

The various validation steps just described provide good evidence for placing each of the structural elements of the inner domain. The N-terminal segment contains Met 65; helices $\alpha1$ and $\alpha5$ and the three-strand sheet are clearly delineated by experimental density; the loop between $\beta5$ and $\beta7$ contains Met 251; the connection between $\beta7$ and the outer domain contains Met 265 and Met 266; and the two ribbons of the bridging sheet appear convincingly in various omit maps. One of these ribbons, the V1-V2 stem, obviously has greater disorder than any of the other structural elements. It also has the highest group thermal parameter (Supplemental Figure S1).

Ordered Glycans

Figure 7 (A and B) summarizes the structures of N-linked oligosaccharides that are ordered in the SIV gp120 core structure. We could detect density for sugars at all 13 glycosylation sites, 7 of which were clearly ($\alpha1$–6) fucosylated. Eight glycans are ordered at least as far as the first branch site. For glycan 459, four residues of one branch were modeled as mannose with ($\alpha1$–2) linkages, but the identity of the last three residues, which make contacts with a symmetry-related molecule, cannot be determined at the current resolution. The long glycans form apparent sugar clusters on the surface of the molecule (Figure 7A). Cluster I includes glycan 277, 305, and 459; of these, glycan 277 also contacts glycan 294 from a symmetry-related molecule, while glycan 459 participates in a crystal contact with polypeptide surface. Cluster II contains glycans 243 and 246; the latter also contacts its counterpart from a symmetry-related molecule. Cluster III contains glycans 376, 475, and 478. Glycans 376 and 478 emerge from a concavity in the gp120 surface of the protein and make several con-

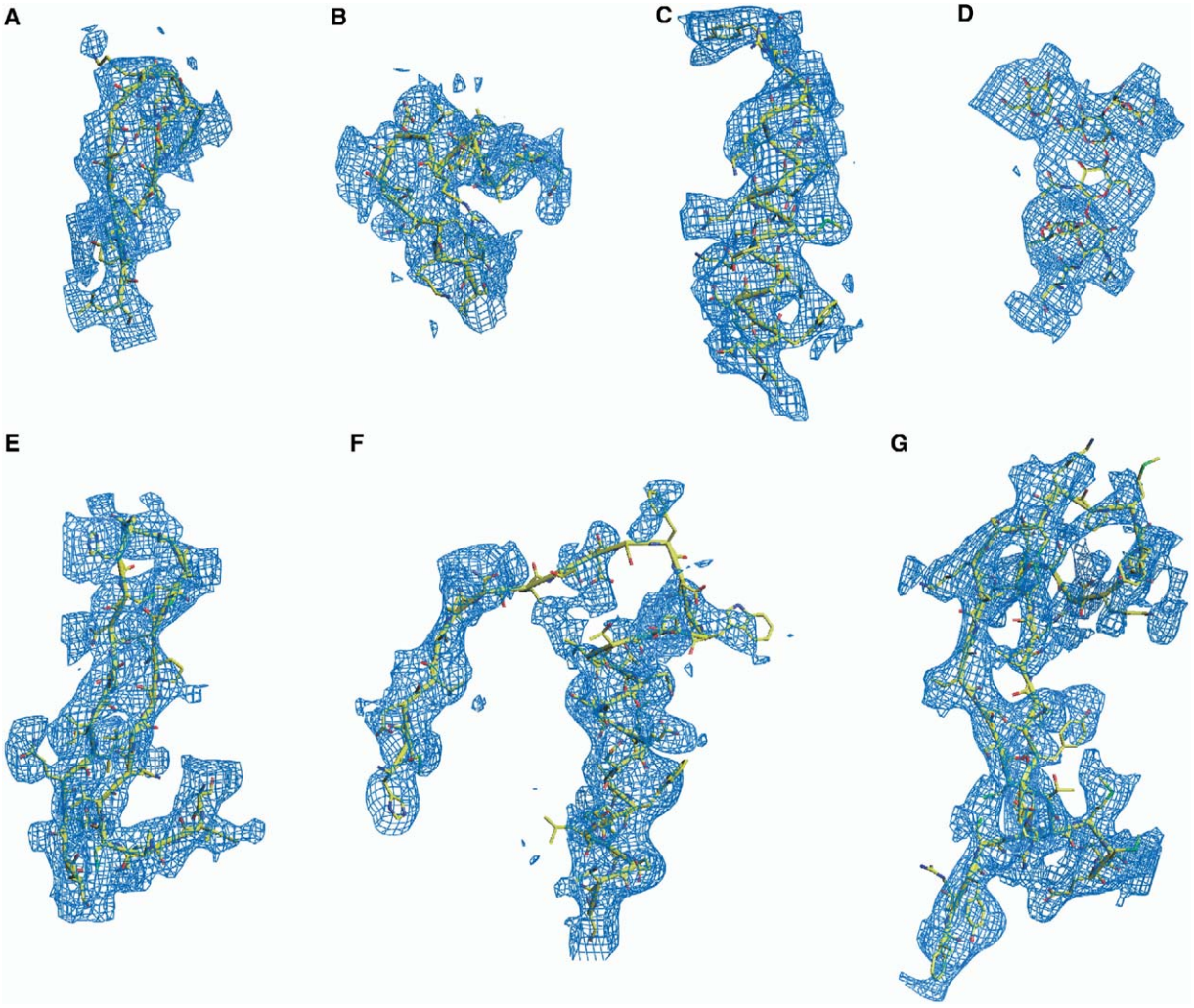

Figure 6. Simulated Annealing Composite Omit Maps Contoured in the Vicinity of Each of the Secondary Structure Elements of the Inner Domain

Maps were calculated using CNS (Brunger et al., 1998) for three different lattices and then averaged, using only one round (no rephasing) of the protocol in Figure 5A. The elements shown in the figure are the hairpin from the outer domain (residues 436–457, A); the short helix and C-terminus (residues 483–499, B); the long helix from the outer domain (residues 347–371, C); glycan 277 (D); the V1-V2 stem (residues 97–219, E); the N terminus and the long helix from the inner domain (residues 64–97, F); the sheet from the inner domain (residues 229–268, G); stick models are in yellow and density in blue. Maps are contoured at 1σ.

tacts with amino-acid side chains. There are presumably hydrogen bond networks within a sugar cluster, as well as bound waters. Therefore, both crystal contacts and interactions with other ordered glycans are likely to contribute to the detectability of particular oligosaccharides in the crystal structure. We note that the iridium sites lie in glycan clusters.

Crystal Contacts

The uncertainty of the side chain conformations at the current resolution makes it difficult to identify with certainty specific interactions, such as Van der Waals contacts, hydrogen bonds, and salt bridges, but there is no ambiguity about which secondary structural elements and glycans contact symmetry-related molecules. The most interesting contact is made by residues near the V4 loop, the base of the V3 loop and glycan 459. The center of this contact is a hydrophobic core, flanked by hydrophilic interactions made by an oligosaccharide and by amino-acid side chains on both molecular surfaces. Near the base of V3, Trp 342, Trp 344, and Tyr 428 from one molecule (molecule I) apparently form a hydrophobic patch, which interacts with the counterpart from a symmetry-related molecule (molecule II). The two branches of glycan 459 from molecule I extend outward to contact both N- and C-terminal segments of V4 from molecule II. The contact continues along the V4 loop, perhaps involving salt bridges between Glu 422 (I) and His 424 (II); it ends at Asn 417 (I), which may hydrogen bond with its symmetry mate (II). This contact, which was invariant in all three lattices we used for multicrystal averaging, creates a crystallo-

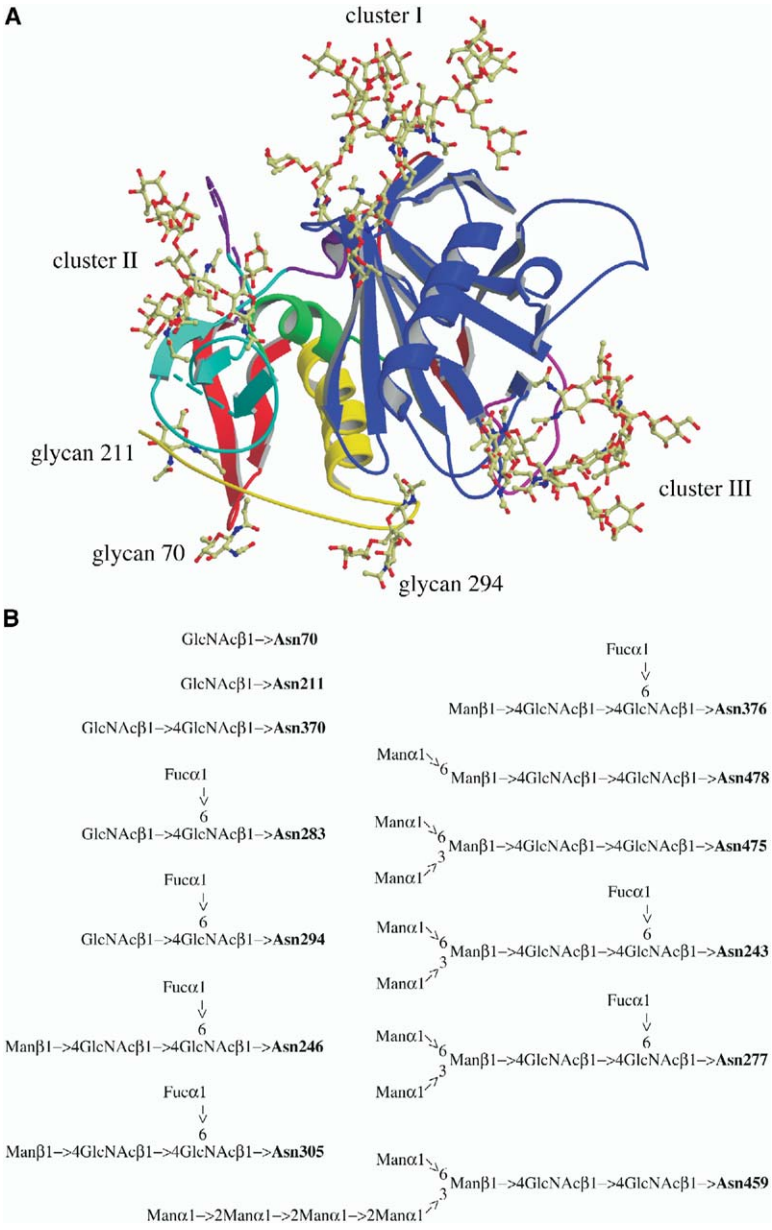

Figure 7. Structures of N-Linked Oligosaccharides Visible in the Final Electron Density Map

Density for all 13 N-linked oligosaccharides was found in the final electron density map.

(A) A view of SIV gp120 core structure in ribbon diagram showing that the long N-linked glycans (stick model) form apparent sugar clusters on the surface of the molecule. Cluster I includes glycan 277, 305, and 459; Cluster II, glycan 243 and 246; Cluster III, glycan 376, 475, and 478.

(B) The structures are shown with ordered residues in abbreviation as well as linkages; GlcNAc, N-acetylglucosamine; Man, mannose; Fuc, fucose; Asn in bold shows the glycosylated asparagine. Glycan 459 was built as a high-mannose type, but the identity of the last three ordered residues of the longer branch, which make contacts with a symmetry-related molecule, is questionable at current resolution.

graphic dyad. The second major contact has on one molecule the base of the hairpin that projects from the outer domain to form part of the bridging sheet in the CD4 bound conformation. On the opposite molecule are the N terminus and part of the inner-domain loop (residues 243 to 255). Two molecules (molecule III and IV), related to each other by a crystallographic dyad, pack against molecule I and II, respectively, through this type of contact; they also make a further contact with each other through a glycan:glycan interaction: glycans 246 from the two molecules pack antiparallel to each other. This molecule I–III contact varies among the three lattices we used for averaging, and the density in this region is affected by the choice of the averaging mask. Additional sugar-sugar interactions are also observed between glycan 277 of molecule I and glycan 294 of molecule IV. Thus, the overall crystal packing involves protein-protein, protein-sugar, and sugar-sugar contacts.

CD4 Binding Induces Major Rearrangements in the gp120 Core

The gp120 of HIV and SIV are closely related. The substantial differences between the unliganded structure described here and the liganded HIV gp120 studied by Kwong et al. (1998) thus reflect primarily the effects of CD4 binding. Figure 8 shows a Cα-trace superimposition of unliganded SIV gp120 core (red) and CD4 bound HIV gp120 core (green). The two outer domains superpose well, except for some surface-exposed loops that have different conformations due to deletions, insertions or different positions of glycosylation sites. The inner domain and bridging sheet, on the other hand, rearrange drastically in response to CD4 binding. Sev-

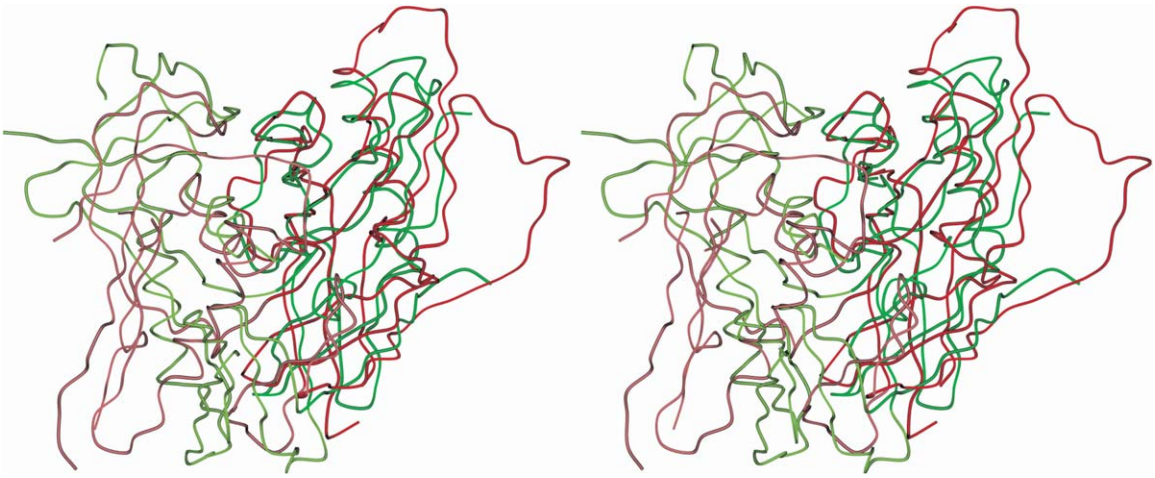

Figure 8. Comparison of Overall Structures of the Unliganded SIV gp120 and the CD4 Bound HIV gp120

A stereo view of superimposed Cα traces for unliganded SIV gp120 core (red) and CD4 bound HIV gp120 core (green). The outer domains of the two proteins, highlighted in dark colors, have essentially the same fold; the inner domains and bridging-sheet elements, in light colors, are markedly different.

eral of the inner-domain secondary-structure elements move independently, rather than in a coordinated fashion. That is, the "inner domain" is not a single, coherent folded structure, but rather a set of distinct structural elements that rearrange with respect to each other. A detailed discussion of the CD4-induced conformational changes will be published elsewhere (Chen et al., in press).

Discussion

Glycoproteins with a modest degree of glycosylation often crystallize readily. For example, a number of variants of influenza haemagglutinin (HA), with 5 to 6 glycans per 60 kDa of protein, have yielded relatively high-resolution structures (Gamblin et al., 2004; Stevens et al., 2004; Wilson et al., 1981). Crystallization can sometimes be impeded, however, by the heterogeneity of glycosylation and the conformational flexibility of carbohydrates. Influenza HA lacks sialic acid, a major source of charge heterogeneity in many glycoproteins from mammalian cells, perhaps contributing to its ease of crystallization. Deglycosylation was essential for obtaining crystals of HIV gp120 core protein suitable for structural studies (Kwong et al., 1999). We were therefore surprised to find that the fully glycosylated SIV gp120 core protein crystallized with 13 N-linked oligosaccharides, when strategies such as deglycosylation or decoration with Fabs had failed. One striking feature of the structure is the extent to which the oligosaccharides form clusters on the protein surface. It is likely that there are extended hydrogen bonding networks within these clusters, perhaps imposing some constraints on the conformation of distal residues in a given glycan. High-resolution data, absent in the current work, would be required to identify such interactions. We note that glycans 70, 211, and 370, which lie outside any of the clusters, have only one or two ordered sugars, consistent with the notion that hydrogen bonding among sugars may limit conformational flexibility. Moreover, some of the glycans participate in crystal contacts, which also appear to be a source of stabilization.

A number of crystal structures have been determined at about 4 Å resolution: examples include the reovirus core (Reinisch et al., 2000), the CD4-MHC class II complex (Wang et al., 2001), the anthrax toxin/receptor complex (Lacy et al., 2004), and the AP1 clathrin adaptor complex (Heldwein et al., 2004). In most of these cases, relatively accurate phases from molecular replacement and high order noncrystallographic symmetry helped in obtaining an interpretable electron density map for model building. Despite the low data-to-parameter ratio, refinement of the resulting models was possible when NCS constraints and strong geometrical restraints were imposed. Brunger and colleagues refined the structure of p97/valosin-containing protein complex with data to only 4.7 Å, aided by a combination of selenomethionine MAD phasing, molecular replacement, and 3-fold NCS (DeLaBarre and Brunger, 2003). In the case of the SIV gp120 core, the initial phases were not very accurate because of the poor phasing power of the heavy atom derivatives and the large differences between the CD4 bound HIV gp120 core and the unliganded SIV gp120 core. Indeed, we initially failed to appreciate the structural differences because of the unexpectedly large number of ordered glycans in our crystals. Moreover, the lack of NCS and the modest solvent content (60%) reduced the power of phase improvement by density modification. Nonetheless, positions of Se atoms in SeMet substituted protein, locations of ordered glycans, use of the CD4 bound HIV gp120 core model as a guide, and multicrystal averaging together enabled us to improve the phases and extend the model gradually.

Refinement was difficult because of a low data-to-parameter ratio. The structure of the AP-1 clathrin adaptor complex, which was built from a far better

map, refined with data to 4 Å only when 6-fold NCS constraints were included in refinement (Heldwein et al., 2004; K.E.H. and S.C.H., unpublished data). It would have been possible in principle to refine the SIV gp120 core structure simultaneously against the three independent data sets from three lattices (Bc44n, Hbc7 and Hbc2/3), to increase the ratio of observations to parameters, but the appropriate refinement algorithm has not yet been implemented. We therefore chose to proceed with refinement against our best native data set by imposing restraints whenever possible, including main chain hydrogen bond restraints and phi-, psi-restraints for helices as well as harmonic restraints for all the glycans. It is useful to set a gradient of harmonic restraints along a glycan with the highest level of restraint assigned to the innermost N-acetyl glucosamine and the lowest to the outermost mannose based on the assumption that distal residues of a given glycan are more mobile than proximal residues. There are exceptions to this assumption–for instance, the distal mannoses of glycan 459 may be fixed by crystal contacts. Refinement against data set Cbc4, which includes data to 3.9 Å, did not yield better results, probably because of incompleteness and anisotropy of the data. Refining the model against both native data and Cbc4 independently helped resolve ambiguities in side chain conformations at later stages of model building.

Efforts to push the resolution of our data beyond the water ring (3.8 Å) will be required to resolve ambiguities and imprecisions in the model. Nonetheless, the model at the current resolution clearly reveals the large rearrangement of the inner domain of gp120 core upon CD4 binding. It is likely to shape our thinking about the organization of the trimeric gp120/gp41 complex on the surface of the virion. We expect that it may also influence strategies for antiviral-drug and AIDS vaccine design.

Experimental Procedures

Protein Production and Crystallization

pSIVgp120core (for definition, see Figure 1) was constructed by standard PCR techniques. We deleted residues from the N- and C termini of SIV gp120, as in Kwong et al. (1999), and we substituted short linkers for the V1-V2 and V3 loops. Two additional residues (His-Met) were introduced by the restriction site (Nde I) at the N terminus. Both restriction digestion and DNA sequencing verified the expression construct. The proteins were expressed using the Bac-to-Bac system (Invitrogen) following a procedure described previously (Chen et al., 2000). Briefly, recombinant baculovirus was generated according to the manufacturer's protocol and amplified in Sf9 insect cells in Hink's TNM-FH medium (JRH Biosciences). The optimal amount of virus and postinfection harvest time were determined by small-scale tests in 100 ml spinner flasks. For large-scale protein production, 12 liter of *T. ni* (Hi-5) cells (2×10^6 cells/ml) in Ex-cell 405 medium (JRH Biosciences) were infected at an MOI of 2.5. The supernatant was harvested 72 hr postinfection by centrifugation and concentrated to 1 liter in a tangential flow filtration system, ProFlux M 12 (Millipore). The concentrated supernatant was clarified by centrifugation.

Expressed SIV gp120 core protein was purified by immunoaffinity chromatography with mAb 17A11 (Edinger et al., 2000; Chen et al., 2000). The concentrated insect cell supernatants were passed through the column at 0.5 ml/min. After extensive washing with PBS, protein was eluted with 100 mM glycine (pH 2.5) and immediately neutralized by 2 M Tris (pH 8.0). Fractions were analyzed by SDS-PAGE. Those containing SIV gp120 core protein were pooled, concentrated, and further purified by gel filtration chromatography on Superdex 200 (Pharmacia) with a buffer containing 25 mM Tris-HCl (pH 8.0) and 150 mM NaCl. The protein was concentrated to 30 $O.D._{280}$/ml, frozen in liquid nitrogen and stored at –80°C. The expected N terminus of the SIV gp120 core protein was confirmed by N-terminal sequencing.

SeMet-substituted SIV gp120 core protein was produced following a published procedure (Carfi et al., 2001). Briefly, 6 L of *T. ni* cells (2×10^6 cells/ml) were infected at an MOI of 2.5. Starvation was initiated by gently spinning the cells (600 g, 5 min) 24 hr postinfection and resuspending them into customized Ex-cell 405 medium from which methionine was omitted (JRH Biosciences). After starvation for 7 hr, the cells were spun and resuspended in customized Ex-cell 405 medium supplemented with 75 mg/L selenomethionine (Sigma) and 5% FBS (Sigma). The supernatant was harvested 72 hr postinfection by centrifugation and concentrated to 0.5 L. SeMet-substituted protein was then purified and stored as described above.

Crystals were grown in hanging drops. The protein was mixed, at 1:1 ratio, with mother liquor containing 15% PEG 6000, 100 mM sodium citrate, (pH 5.0) and 8% PEG 400 at 20°C. Crystals appeared in a few days and grew to about 100 microns in the largest dimension in 2–3 weeks. 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, and flash-freezing in liquid nitrogen. Smaller crystals (about 70 microns) were obtained with SeMet-substituted SIV gp120 core protein under the same conditions. Amino acid analysis of the SeMet-substituted protein showed an incorporation rate of at least 50%. To make heavy atom derivatives, over 60 heavy atom compounds were dissolved in soaking solution containing 15% PEG 6000, 100 mM sodium citrate, (pH 5.0) and 8% PEG 400. The final concentration was adjusted to 1 mM when possible, and a saturated solution was used if the compound was not soluble at 1 mM. Soaks were allowed to proceed for 24 hr at 20°C.

Data Collection and Phase Improvement

Data were collected with 0.5 or 1° oscillation and an exposure time of 3 min per frame at CHESS F1, or 5 s per frame at APS 19ID (where no beam attenuation was used). To use synchrotron beamtime more efficiently, crystals soaked with heavy atom compounds were screened by collecting a 20-frame "mini data set," which was scaled to the native data set. Complete data were collected only for derivatives that showed promising statistics after mini data set scaling. Data were also collected for K_2IrCl_6 soaked crystals in an inverse beam mode at beamline F1 to maximize anomalous signal (Ibc1 in Table1), or with larger beam stop-to-crystal distance to include low-resolution reflections (Sbc10L in Table 1). A single-wavelength (peak) data set from SeMet crystals, collected at APS 19ID, extended to about 5 Å; data collected at other wavelengths (inflection and remote) were not useful because of deterioration from radiation damage. Denzo and Scalepack or HKL2000 were used for data processing (Otwinowski and Minor, 1997). Molecular replacement was performed using the program AMORE (Navaza and Saludjian, 1997)

Subsequent phase improvement by iterative cycles of phase combination (partial model phases and experimental phases), density modification, and data sharpening were performed using the CCP4 programs SIGMAA, DM, and CAD (Collaborative Computational Project, 1994). Initially, data were sharpened by applying a B-factor of –200 to the amplitudes in order to show side chain density. With improvement of the model phases, the sharpening B-factor could be reduced to –100. Transformation matrices relating model positions in three different crystal lattices—native (Bc44n), $IrCl_3$-soaked (Hbc7), and CdI-soaked (Hbc2/3) crystals—were determined from both rigid-body refinement using CNS and molecular replacement using AMORE. Multicrystal averaging was then performed using a C-shell script to run the CCP4 programs MAPROT for map averaging, RSTATS for data scaling, and FFT for map calculation (Figure 5).

Model Building and Refinement

Program O was used for model building (Jones et al., 1991) and both CNS (Brunger et al., 1998) and Refmac 5 (Murshudov, 1997),

for heavily restrained refinement. Torsional simulated annealing and minimization in CNS were carried out initially with the phase-restrained MLHL target and later with the MLF target. TLS refinement, followed by limited minimization with a flat bulk solvent model, was performed with Refmac 5, as described in the main text. The final model contains all residues in the gp120 core, plus the vector-derived His64 and Met65, with the exception of residues 220–228 and 500–502.

Supplemental Data

Supplemental data include three figures and one table and can be found with this article online at http://www.structure.org/cgi/content/full/13/2/197/DC1/.

Accession Numbers

The coordinates have been deposited in the Protein Data Bank for accession number 2BF1.

Acknowledgments

We thank staff at CHESS beamline F1 and APS beamline 19ID for assistance, Dr. James Hoxie of University of Penn, for hybridomas and members of 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 Innovation Grant Program for Approaches in HIV Vaccine Research (to S.C.H. and D.C.W.), and by a HIVRAD grant AI43649 (to Ellis Reinherz). S.C.H. is, and D.C.W. was, an Investigator of the Howard Hughes Medical Institute.

Received: October 14, 2004
Revised: December 3, 2004
Accepted: December 6, 2004
Published: February 8, 2005

References

Allan, J.S., Coligan, J.E., Barin, F., McLane, M.F., Sodroski, J.G., Rosen, C.A., Haseltine, W.A., Lee, T.H., and Essex, M. (1985). Major glycoprotein antigens that induce antibodies in AIDS patients are encoded by HTLV-III. Science *228*, 1091–1094.

Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., et al. (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr. D Biol. Crystallogr. *54*, 905–921.

Burton, D.R., Desrosiers, R.C., Doms, R.W., Koff, W.C., Kwong, P.D., Moore, J.P., Nabel, G.J., Sodroski, J., Wilson, I.A., and Wyatt, R.T. (2004). HIV vaccine design and the neutralizing antibody problem. Nat. Immunol. *5*, 233–236.

Carfi, A., Willis, S.H., Whitbeck, J.C., Krummenacher, C., Cohen, G.H., Eisenberg, R.J., and Wiley, D.C. (2001). Herpes simplex virus glycoprotein D bound to the human receptor HveA. Mol. Cell *8*, 169–179.

Center, R.J., Schuck, P., Leapman, R.D., Arthur, L.O., Earl, P.L., Moss, B., and Lebowitz, J. (2001). Oligomeric structure of virion-associated and soluble forms of the simian immunodeficiency virus envelope protein in the prefusion activated conformation. Proc. Natl. Acad. Sci. USA *98*, 14877–14882.

Center, R.J., Leapman, R.D., Lebowitz, J., Arthur, L.O., Earl, P.L., and Moss, B. (2002). Oligomeric structure of the human immunodeficiency virus type 1 envelope protein on the virion surface. J. Virol. *76*, 7863–7867.

Chan, D.C., Fass, D., Berger, J.M., and Kim, P.S. (1997). Core structure of gp41 from the HIV envelope glycoprotein. Cell *89*, 263–273.

Chen, B., Zhou, G., Kim, M., Chishti, Y., Hussey, R.E., Ely, B., Skehel, J.J., Reinherz, E.L., Harrison, S.C., and Wiley, D.C. (2000). Expression, purification, and characterization of gp160e, the soluble, trimeric ectodomain of the simian immunodeficiency virus envelope glycoprotein, gp160. J. Biol. Chem. *275*, 34946–34953.

Chen, B., Vogan, E.M., Gong, H., Skehel, J.J., Wiley, D.C., and Harrison, S.C. (2005). Structure of an unliganded simian immunodeficiency virus gp120 core. Nature, in press.

Collaborative Computational Project, N. (1994). The CCP4 suite: programs for protein crystallography. Acta Crystallogr. D Biol. Crystallogr. *50*, 760–763.

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

de la Fortelle, E., and Bricogne, G. (1997). Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. Methods Enzymol. *276*, 472–493.

DeLaBarre, B., and Brunger, A.T. (2003). Complete structure of p97/valosin-containing protein reveals communication between nucleotide domains. Nat. Struct. Biol. *10*, 856–863.

Edinger, A.L., Ahuja, M., Sung, T., Baxter, K.C., Haggarty, B., Doms, R.W., and Hoxie, J.A. (2000). Characterization and epitope mapping of neutralizing monoclonal antibodies produced by immunization with oligomeric simian immunodeficiency virus envelope protein. J. Virol. *74*, 7922–7935.

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

Gamblin, S.J., Haire, L.F., Russell, R.J., Stevens, D.J., Xiao, B., Ha, Y., Vasisht, N., Steinhauer, D.A., Daniels, R.S., Elliot, A., et al. (2004). The structure and receptor binding properties of the 1918 influenza hemagglutinin. Science *303*, 1838–1842.

Heldwein, E.E., Macia, E., Wang, J., Yin, H.L., Kirchhausen, T., and Harrison, S.C. (2004). Crystal structure of the clathrin adaptor protein 1 core. Proc. Natl. Acad. Sci. USA *101*, 14108–14113.

Hsu, T.A., Takahashi, N., Tsukamoto, Y., Kato, K., Shimada, I., Masuda, K., Whiteley, E.M., Fan, J.Q., Lee, Y.C., and Betenbaugh, M.J. (1997). Differential N-glycan patterns of secreted and intracellular IgG produced in Trichoplusia ni cells. J. Biol. Chem. *272*, 9062–9070.

Jones, T.A., Zou, J.Y., and Cowan, S.W. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr. A *47*, 110–119.

Kilby, J.M., and Eron, J.J. (2003). Novel therapies based on mechanisms of HIV-1 cell entry. N. Engl. J. Med. *348*, 2228–2238.

Kim, M., Chen, B., Hussey, R.E., Chishti, Y., Montefiori, D., Hoxie, J.A., Byron, O., Campbell, G., Harrison, S.C., and Reinherz, E.L. (2001). The stoichiometry of trimeric SIV glycoprotein interaction with CD4 differs from that of anti-envelope antibody Fab fragments. J. Biol. Chem. *276*, 42667–42676.

Kwong, P.D., Wyatt, R., Robinson, J., Sweet, R.W., Sodroski, J., and Hendrickson, W.A. (1998). Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. Nature *393*, 648–659.

Kwong, P.D., Wyatt, R., Desjardins, E., Robinson, J., Culp, J.S., Hellmig, B.D., Sweet, R.W., Sodroski, J., and Hendrickson, W.A. (1999). 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.

Kwong, P.D., Wyatt, R., Majeed, S., Robinson, J., Sweet, R.W., Sodroski, J., and Hendrickson, W.A. (2000). Structures of HIV-1 gp120 envelope glycoproteins from laboratory-adapted and primary isolates. Struct. Fold. Des. *8*, 1329–1339.

Lacy, D.B., Wigelsworth, D.J., Melnyk, R.A., Harrison, S.C., and Collier, R.J. (2004). Structure of heptameric protective antigen bound to an anthrax toxin receptor: a role for receptor in pH-dependent pore formation. Proc. Natl. Acad. Sci. USA *101*, 13147–13151.

Lin, P.F., Blair, W., Wang, T., Spicer, T., Guo, Q., Zhou, N., Gong, Y.F., Wang, H.G., Rose, R., Yamanaka, G., et al. (2003). 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.

Moore, J.P., McKeating, J.A., Weiss, R.A., and Sattentau, Q.J. (1990). Dissociation of gp120 from HIV-1 virions induced by soluble CD4. Science *250*, 1139–1142.

Murshudov, G.N. (1997). Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr. D Biol. Crystallogr. *53*, 240–255.

Navaza, J., and Saludjian, P. (1997). AMoRe: an automated molecular replacement program package. Methods Enzymol. *276*, 581–594.

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. *276*, 307–326.

Ravelli, R.B., and McSweeney, S.M. (2000). The 'fingerprint' that X-rays can leave on structures. Struct. Fold. Des. *8*, 315–328.

Reinisch, K.M., Nibert, M.L., and Harrison, S.C. (2000). Structure of the reovirus core at 3.6 Å resolution. Nature *404*, 960–967.

Rizzuto, C., and Sodroski, J. (2000). Fine definition of a conserved CCR5-binding region on the human immunodeficiency virus type 1 glycoprotein 120. AIDS Res. Hum. Retroviruses *16*, 741–749.

Rizzuto, C.D., Wyatt, R., Hernández-Ramos, N., Sun, Y., Kwong, P.D., Hendrickson, W.A., and Sodroski, J. (1998). A conserved HIV gp120 glycoprotein structure involved in chemokine receptor binding. Science *280*, 1949–1953.

Sattentau, Q.J., and Moore, J.P. (1991). Conformational changes induced in the human immunodeficiency virus envelope glycoprotein by soluble CD4 binding. J. Exp. Med. *174*, 407–415.

Sattentau, Q.J., Moore, J.P., Vignaux, F., Traincard, F., and Poignard, P. (1993). Conformational changes induced in the envelope glycoproteins of the human and simian immunodeficiency viruses by soluble receptor binding. J. Virol. *67*, 7383–7393.

Starcich, B.R., Hahn, B.H., Shaw, G.M., McNeely, P.D., Modrow, S., Wolf, H., Parks, E.S., Parks, W.P., Josephs, S.F., Gallo, R.C., et al. (1986). Identification and characterization of conserved and variable regions in the envelope gene of HTLV-III/LAV, the retrovirus of AIDS. Cell *45*, 637–648.

Stein, P.E., Boodhoo, A., Armstrong, G.D., Cockle, S.A., Klein, M.H., and Read, R.J. (1994). The crystal structure of pertussis toxin. Structure *2*, 45–57.

Stevens, J., Corper, A.L., Basler, C.F., Taubenberger, J.K., Palese, P., and Wilson, I.A. (2004). Structure of the uncleaved human H1 hemagglutinin from the extinct 1918 influenza virus. Science *303*, 1866–1870.

Terwilliger, T.C., and Berendzen, J. (1999). Automated MAD and MIR structure solution. Acta Crystallogr. D Biol. Crystallogr. *55*, 849–861.

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

Veronese, F.D., DeVico, A.L., Copeland, T.D., Oroszlan, S., Gallo, R.C., and Sarngadharan, M.G. (1985). Characterization of gp41 as the transmembrane protein coded by the HTLV-III/LAV envelope gene. Science *229*, 1402–1405.

Wang, J.H., Meijers, R., Xiong, Y., Liu, J.H., Sakihama, T., Zhang, R., Joachimiak, A., and Reinherz, E.L. (2001). Crystal structure of the human CD4 N-terminal two-domain fragment complexed to a class II MHC molecule. Proc. Natl. Acad. Sci. USA *98*, 10799–10804.

Wang, T., Zhang, Z., Wallace, O.B., Deshpande, M., Fang, H., Yang, Z., Zadjura, L.M., Tweedie, D.L., Huang, S., Zhao, F., et al. (2003). Discovery of 4-benzoyl-1-[(4-methoxy-1H-pyrrolo[2,3-b]pyridin-3-yl)oxoacetyl]-2-(R)-methylpiperazine (BMS-378806): a novel HIV-1 attachment inhibitor that interferes with CD4-gp120 interactions. J. Med. Chem. *46*, 4236–4239.

Wei, X., Decker, J.M., Wang, S., Hui, H., Kappes, J.C., Wu, X., Salazar-Gonzalez, J.F., Salazar, M.G., Kilby, J.M., Saag, M.S., et al. (2003). Antibody neutralization and escape by HIV-1. Nature *422*, 307–312.

Weissenhorn, W., Dessen, A., Harrison, S.C., Skehel, J.J., and Wiley, D.C. (1997). Atomic structure of the ectodomain from HIV-1 gp41. Nature *387*, 426–430.

Weissenhorn, W., Dessen, A., Calder, L.J., Harrison, S.C., Skehel, J.J., and Wiley, D.C. (1999). Structural basis for membrane fusion by enveloped viruses. Mol. Membr. Biol. *16*, 3–9.

Wild, C.T., Shugars, D.C., Greenwell, T.K., McDanal, C.B., and Matthews, T.J. (1994). Peptides corresponding to a predictive alpha-helical domain of human immunodeficiency virus type 1 gp41 are potent inhibitors of virus infection. Proc. Natl. Acad. Sci. USA *91*, 9770–9774.

Wilson, I.A., Skehel, J.J., and Wiley, D.C. (1981). Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 A resolution. Nature *289*, 366–373.

Wu, L., Gerard, N.P., Wyatt, R., Choe, H., Parolin, C., Ruffing, N., Borsetti, A., Cardoso, A.A., Desjardin, E., Newman, W., et al. (1996). CD4-induced interaction of primary HIV-1 gp120 glycoproteins with the chemokine receptor CCR-5. Nature *384*, 179–183.

Wyatt, R., Sullivan, N., Thali, M., Repke, H., Ho, D., Robinson, J., Posner, M., and Sodroski, J. (1993). Functional and immunologic characterization of human immunodeficiency virus type 1 envelope glycoproteins containing deletions of the major variable regions. J. Virol. *67*, 4557–4565.

Wyatt, R., Kwong, P.D., Desjardins, E., Sweet, R.W., Robinson, J., Hendrickson, W.A., and Sodroski, J.G. (1998). The antigenic structure of the HIV gp120 envelope glycoprotein. Nature *393*, 705–711.