Protein Science (1997), 6:1653–1660. Cambridge University Press. Printed in the USA.

The extracellular domain of immunodeficiency virus gp41 protein: Expression in *Escherichia coli*, purification, and crystallization

PAUL T. WINGFIELD,[1] STEPHEN J. STAHL,[1] JOSHUA KAUFMAN,[1] ADAM ZLOTNICK,[1] C. CRAIG HYDE,[2] ANGELA M. GRONENBORN,[3] AND G. MARIUS CLORE[3]

[1]Protein Expression Laboratory, National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health, Bethesda, Maryland 20892-2775
[2]Laboratory of Structural Biology, National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health, Bethesda, Maryland 20892-2775
[3]Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892-0520

(RECEIVED March 18, 1997; ACCEPTED April 24, 1997)

Abstract

The *env* gene of SIV and HIV-1 encodes a single glycoprotein gp160, which is processed to give a noncovalent complex of the soluble glycoprotein gp120 and the transmembrane glycoprotein gp41. The extracellular region (ectodomain), minus the N-terminal fusion peptide, of gp41 from HIV-1 (residues 27–154) and SIV (residues 27–149) have been expressed in *Escherichia coli*. These insoluble proteins were solubilized and subjected to a simple purification and folding scheme, which results in high yields of soluble protein. Purified proteins have a trimeric subunit composition and high α-helical content, consistent with the predicted coil-coil structure. SIV gp41 containing a double cysteine mutation was crystallized. The crystals are suitable for X-ray structure determination and, preliminary analysis, together with additional biochemical evidence, indicates that the gp41 trimer is arranged as a parallel bundle with threefold symmetry.

Keywords: *E. coli* protein expression; HIV-1 and SIV gp41 ectodomain; physiochemical analysis; protein crystallization; protein folding and purification; X-ray diffraction

The envelope glycoprotein of HIV-1 is synthesized as a large glycosylated precursor (gp160) that is processed proteolytically to the external glycoprotein gp120 and the smaller transmembrane glycoprotein gp41 (Capon & Ward, 1991; Geleziunas et al., 1994). The mature proteins are noncovalently associated. Gp41 anchors the glycoprotein complex to the viral membrane, or plasma membrane, and carries a hydrophobic fusion peptide at its N terminus (Bosch et al., 1989). Interaction of gp120, expressed either on a viral particle or on the surface of infected cells, with the HIV specific cell-surface receptor CD4 (Bour et al., 1995) results in a conformational change and possibly proteolytic cleavage of regions of the HIV envelope not involved directly in CD4 binding (Kowalski et al., 1991). As a result of the gp120-induced change, gp41 exerts its fusogenic capacity, either by interacting with an as yet unidentified second receptor on the target, cell or by direct insertion in the target cell membrane.

The N-terminal region of gp41 exposed on the outer surface of the viral membrane is called the ectodomain and portions of this ~150 residue sequence from both HIV and SIV have been expressed recently in *Escherichia coli* (Blacklow et al., 1995; Lu et al., 1995). Protein expressed in *E. coli* was digested proteolytically into two small subdomains comprising residues 28–80 (N-domain) and 107–149 (C-domain), eliminating a cysteine-containing sequence. When the N and C domains were mixed, they formed a stable helical hexameric complex consisting of three sets of the N+C peptides. Based on these findings, these authors proposed a model for the gp41 ectodomain of a parallel trimeric coiled coil-coil of the N-domain encircled by three C-domains with the N- and C-domain helices arranged antiparallel to one another. A soluble form of the HIV ectodomain (residues 21–166) produced in insect cells (Weissenhorn et al., 1996) was also an oligomeric protein with a high α-helical content. Electron microscopy of this glycoprotein indicated a rod-like shape, which the authors propose is in the membrane fusion-active, extended structure, analogous to the influenza virus hemagglutinin in its low-pH-induced conformation (Bullough et al., 1994).

Reprint requests to: P.T. Wingfield, National Institutes of Health, Building 6B, Room 1B130, 6 Center Dr. MSC 2775, Bethesda, Maryland 20892-2775; e-mail: pelpw@helix.nih.gov.

Abbreviations: HIV gp41$^{27-154}$, residues 538–665 of the HIV-1 HXB2 *env* gene; MPD, 2-methyl-2,4 pentanediol; SIV gp41$^{27-149}$, residues 552–674 of the SIV$_{mac}$239 *env* gene.

In this paper, we report on the expression of the SIV gp41 residues 27–149 (SIV gp41$^{27-149}$) and the HIV gp41 residues 27–154 (HIV gp41$^{27-154}$) in *E. coli*. The regions expressed do not contain the N-terminal fusion domain. A simple purification and folding scheme is described that takes advantage of the acid stability of the gp41 proteins. Characterization by CD and analytical ultracentrifugation indicate proteins with a secondary structure consisting of >80% α-helix and a trimeric subunit composition. In both the HIV and SIV proteins, the two cysteines in a putative loop region were substituted by alanine residues in order to enhance the physical properties of the proteins. Crystallization of the SIV gp41$^{27-149}$ with the double cysteine to alanine mutation into diamond-shaped plates is described. Preliminary X-ray diffraction analysis, together with other biochemical evidence, indicate that the SIV gp41$^{27-149}$ trimer is a parallel bundle.

Results

Protein purification and folding

SIV gp41$^{27-149}$ and HIV gp41$^{27-154}$ accumulated as insoluble proteins when expressed in *E. coli*. The expression level of the SIV protein was extremely high (Fig. 2, lane B) and similar in both complex and minimal media. HIV gp41$^{27-154}$ was expressed at about 10–20% of the level of the SIV protein (data not shown). Insoluble protein was extracted with guanidine-HCl and purified by one step of gel filtration, also in the presence of guanidine-HCl (Fig. 1). Analysis of protein in the main peak (labeled 6 in Fig. 1) by analytical ultracentrifugation and CD indicated that it was monomeric and unfolded (data not shown). Protein in peak 6 was used for further processing, and other protein, which was either folded partially or aggregated (e.g., peaks 1–4 in Fig. 1), was discarded. Prior to folding the protein, the guanidine-HCl was exchanged by reverse phase (RP)-HPLC for acetonitrile-TFA. Although this step provided little in the way of extra purification, we observed that folding protein from this solvent was more efficient (higher yielding) than folding directly from guanidine-HCl. Protein was folded by dialysis against a weakly acidic, pH 3.0, buffer. At this pH, the protein assumed a native-like conformation (see below for detail) and had much a higher solubility than at higher pH values, for example, greater than 4.5. We also observed that the solubility of the protein was higher at low ionic strengths; hence, salt was omitted from the folding buffer.

After concentration of the folded protein, a small amount of aggregated protein formed that was removed by gel filtration in the presence of 10% (v/v) 2-propanol. It was observed that, in order to gel filtrate moderately concentrated protein (>5 mg/mL) on either silica-based or dextran-based (e.g., Superdex) matrices, a solvent additive was required to prevent the protein from sticking (adsorbing) to the column. 2-Propanol (10–20%) or 200 mM guanidine-HCl were both useful in this respect and neither additive perturbed the overall tertiary structure of the protein as evidenced by near-UV CD (see below).

Using minimal media for fermentation, we obtained about 15 g (wet weight) of cells/L and, from complex media, greater than 60 g/L of cells. In either case, more than 20 mg of purified protein in guanidine-HCl per gram of cells was recovered. After protein folding and the gel filtration stage, 70–80% of the denatured starting materials was obtained as native protein. Hence, a typical overall recovery of SIV gp41$^{27-149}$ from 15 g of cells is about 200 mg of protein. About 20–30 mg of the HIV gp41$^{27-154}$ was recovered from the same weight of cells for both the HIV- or SIV-derived proteins; the introduction of the double cysteine mutation had no effect on either expression or recovery yields.

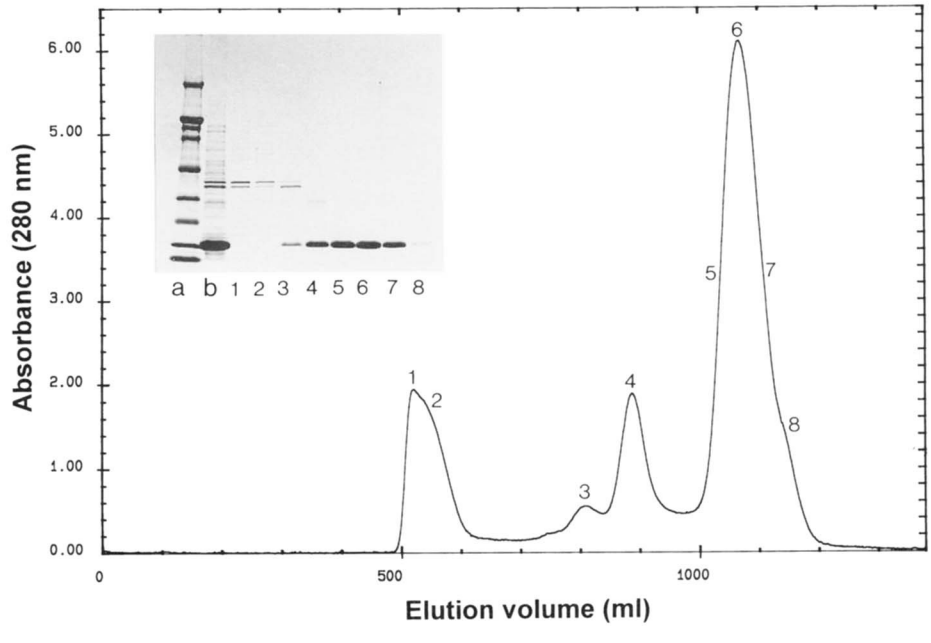

Fig. 1. Gel filtration of SIV gp41$^{27-149}$ in guanidine-HCl. Insoluble protein was solubilized from *E. coli* lysates with 8 M guanidine-HCl and applied to a column of Superdex 200 equilibrated in 4 M guanidine-HCl. SDS-PAGE of the numbered fractions is shown in the insert. Lane a, molecular weight standards, the first five bands of which (bottom to top) correspond to: 6, 14.4, 21.5, 31, 45, and 66.2 kDa. Lane b, the starting material applied to the column. Protein in the main peak (fractions 5–7) was used for protein folding.

Physicochemical characterization

N-terminal sequencing of both the wild-type and mutant SIV proteins indicated that, in both cases, about 95% of the initiating N-formylmethionine had been processed and the N-terminus was alanine, corresponding to residue 27 of the native sequence. MALDI-TOF mass spectrometry of the SIV wild-type and mutant proteins (data not shown) gave mass values within 0.05% of that predicted from the respective coding DNA sequences (i.e., 14,480 and 14,416). SDS-PAGE of the purified proteins showed single bands with molecular weights of about 14,000–15,000 (Fig. 2, lanes C and D; data for HIV protein not shown). Gel analysis was also performed on samples preincubated for several days without reductant present in the buffer, then treated with SDS-PAGE sample buffer also without reductant. As expected, the SIV mutant protein gave the same result as reduced protein (Fig. 2, compare lanes D and F), whereas the wild-type protein now exhibited two additional bands (Fig. 2, lanes G and H) corresponding in molecular weights to dimer and trimer. Quenching free sulfhydryl with iodoacetamide prior to addition of the SDS-sample buffer resulted in a more intense trimer band (Fig. 2G). This indicates that sulfhydryl groups in the wild-type protein are capable of forming intermolecular crosslinks with neighboring subunits (discussed further below).

Native molecular weights were determined by sedimentation equilibrium measurements. Typically at pH 3.5, SIV gp41$^{27-149}$ in the presence of reductant [5 mM dithiothreitol (DTT)] has a molecular weight of about 38,000–39,000, indicating that, under these conditions, the protein is a trimer (mass predicted from DNA coding sequence: 43,440). When reductant is omitted from the buffer, a higher average mass of about 45,000 is observed; it is also apparent that there is a small amount of aggregation, which increases over time. In contrast, the SIV gp41$^{27-149}$ with the double cysteine mutation in the absence of reductant gave an average mass of about 36,000–38,000, i.e., slightly lower than predicted for a stable trimer (43,249), but similar to the reduced wild-type protein. The data for the mutant and wild-type protein, the latter in the presence of reductant, were best modeled by assuming a reversible monomer–trimer association with a determined self association constant (K_a) of $\sim 1.5 \times 10^{11}$ M^{-2}. This can be compared with a K_a of 4.8×10^{11} M^{-2} determined for the analogous HIV gp41$^{27-154}$ cysteine mutant. In practical terms, these association constants predict, for example, that at 1 mg/mL (69 μM monomer) ~7% (SIV) and ~5% (HIV) of the gp41 proteins are monomeric.

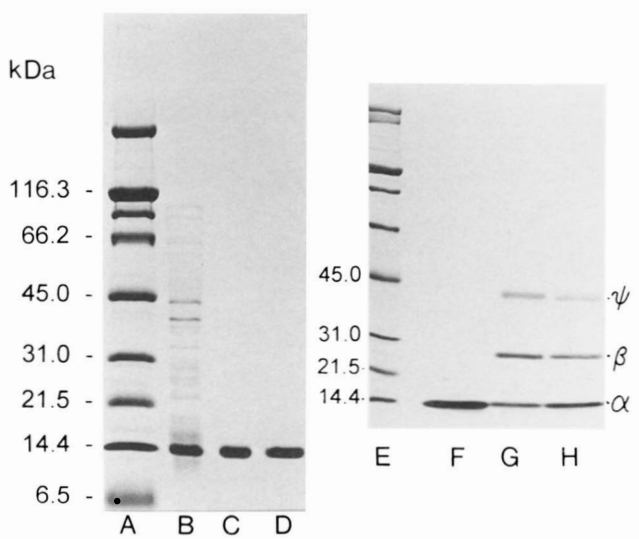

Fig. 2. SDS-PAGE of SIV gp41$^{27-149}$. Lane A, molecular weight standards the masses of which, in kDa, are indicated; lane B, *E. coli* lysate from cells grown in minimal media; lane C, wild-type protein; lane D, double mutant C86A, C92A, lane E, molecular weight standards. Lanes F, G, and H refer to proteins analyzed under nonreducing conditions (no DTT in the SDS-gel sample buffer) following either alkylation (ANR) or not (NR) with 20 mM iodoacetamide (to cap free sulfhydryl groups). Lane F, double mutant (NR); lane G, wild-type protein (ANR); lane H, as for G (NR). Positions of monomeric (α), dimeric (β), and trimeric protein (γ) are indicated.

Conformation and shape

The overall conformation was examined by CD. For the SIV wild-type and double cysteine mutant proteins, the similarity in the far-UV and the near-UV (conformational fingerprint region) suggests that the two proteins have similar overall secondary and tertiary conformations. Using the far-UV circular dichroic spectra (Fig. 3A), it was estimated, using several methods (Chang et al., 1978; Perczel et al., 1992; Sreerama & Woody, 1993), that both the wild-type and mutant proteins are constituted of about 80% α-helix. It has been observed with other proteins containing helical coil-coil motifs (see, for example, Monera et al., 1994), that the far-UV spectrum is not very sensitive to changes in tertiary structure. For example, the SIV gp41$^{27-149}$ double cysteine mutant in 20% acetonitrile, 0.1% TFA (solvent used for reverse phase chromatography), appears fairly well structured in the far-UV region, with only a small decrease in the ellipticity at 220 nm compared with that at 209 nm (Fig. 3A, spectrum 3). Contrast this with the complete loss of native-like signal in the near-UV region (Fig. 3B, spectrum 3). We examined the conformation of the protein over a fairly wide range of pH (3–9.5) in aqueous buffers and saw little change in the spectral properties. Although the mildly acidic solvent used for folding and analysis may be considered nonphysiological, the proteins obviously exhibit acid stability and, as mentioned above, are much more soluble at low pH.

In the far-UV region, the CD spectra of the HIV and SIV proteins are very similar, indicating similar α-helical secondary structure (compare Fig. 3A, spectrum 2 and Fig. 3C). In the near-UV region, both proteins exhibit a main peak of negative ellipticity at around 292 nm; the HIV protein has a second negative peak at 285–286 nm (Fig. 3D), whereas the SIV protein has a weak shoulder at this wavelength (Fig. 3B, spectrum 2). Other features of the two spectra are similar. In the HIV and SIV proteins, there are six conserved Trp residues, HIV gp41$^{27-154}$ has one additional Trp at position 103, which must account for the spectral differences in otherwise apparently similarly folded proteins.

Information on the shape of the protein can be obtained by sedimentation velocity analysis because the sedimentation coefficient (s) is shape dependent (Van Holde, 1975). If we consider gp41 to be a simple sphere of molecular weight about 44,000, with a typical hydration value of 0.35 g water per g protein, it would be expected to have an s value of about 3.8. The observed s values for the SIV and HIV proteins of 2.8 and 3.0 are significantly lower, and indicate extended or rod-like structures (see the Discussion).

Crystallographic characterization of gp41

Crystals of the mutant gp41, grown from MPD (Fig. 4), diffracted to 3.8 Å. Both R-axis (Higashi, 1990) and DENZO (Otwinowski, 1993) autoindexing routines identified the space group as C2. After

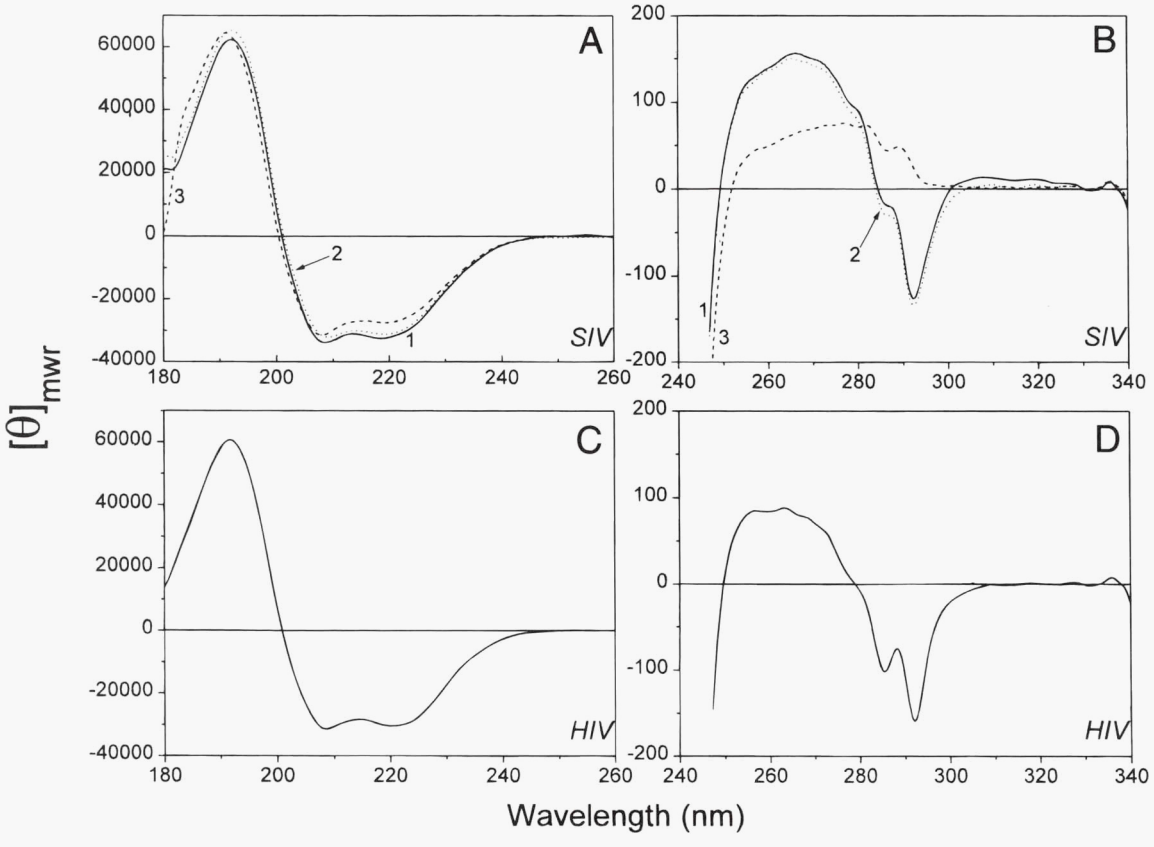

Fig. 3. Circular dichroic spectra of SIV and HIV gp41 ectodomains. **A,B:** SIV gp41$^{27-149}$. Spectra 1 (solid), wild-type protein; spectra 2 (dotted), double mutant C86A, C92A in 50 mM sodium format, pH 3.5; spectra 3 (dashed), double mutant in 20% acetonitrile, 0.1% TFA, pH 1.9. **C,D:** HIV gp41$^{27-154}$ C87A, C93A in 50 mM sodium formate, pH 3.5. Spectra are shown in far-UV region to limits of reliable detection as monitored by photomultiplier voltage (usually less than 600 V). Units of the ordinates are mean residue ellipticity $[\theta]_{MWR}$ and have the dimension, deg·cm^2·dmol^{-1}.

postrefinement of a partial data set, unit cell dimensions were a = 94.1 Å, b = 50.5 Å, c = 75.1 Å, $\alpha = \gamma = 90°$, β = 100.7°. In space group C2, the unit cell is divided into four asymmetric units related by twofold symmetry and 2_1 screw axes, where all symmetry axes are parallel to the unit cell b axis (Fig. 5B). The volume per Da of protein, the Matthews coefficient (V_m), is 2.14 Å^3/Da when calculated for a trimer of SIV gp41$^{27-149}$ in the asymmetric unit. This value is in the center of the typical range for globular proteins (Matthews, 1968).

Rotation functions (Rossmann, 1990) were calculated to characterize the molecular symmetry of the asymmetric unit and to determine the orientation of the symmetry axis within the unit cell. Although only a partial data set was collected (Table 1), the rotation function calculation was sufficiently robust to yield an unambiguous result. A search for threefold symmetry, conducted in polar coordinates, revealed a single peak that was consistent for several ranges of resolution (see Tong & Rossmann, 1990). The peak shown in Figure 5A is 11.9 σ above background. A fine search identifies the coordinates of this peak as ϕ = 207.4°, ψ = 90°, and κ = 120°. Thus, the threefold axis is perpendicular to the crystallographic twofold b axis and is 62.6° from the crystallographic c axis (Fig. 5B). Because the noncrystallographic symmetry axes are perpendicular to the crystallographic twofold axis, crystallographic symmetry requires that all four molecules in the

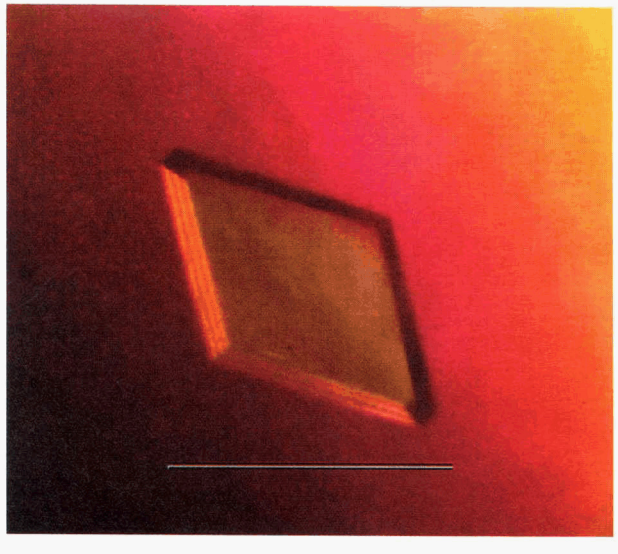

Fig. 4. Crystal of SIV gp41$^{27-149}$. Crystals typically grow in diamond-shaped plates. This crystal grew in 4 weeks at 23 °C in a sitting drop, vapor-diffusion experiment as described in the text, containing 27% (v/v) 2-propanol, 50 mM sodium acetate, pH 5.5, and 0.2 M $CaCl_2$, in the reservoir. Scale bar is 0.4 mm.

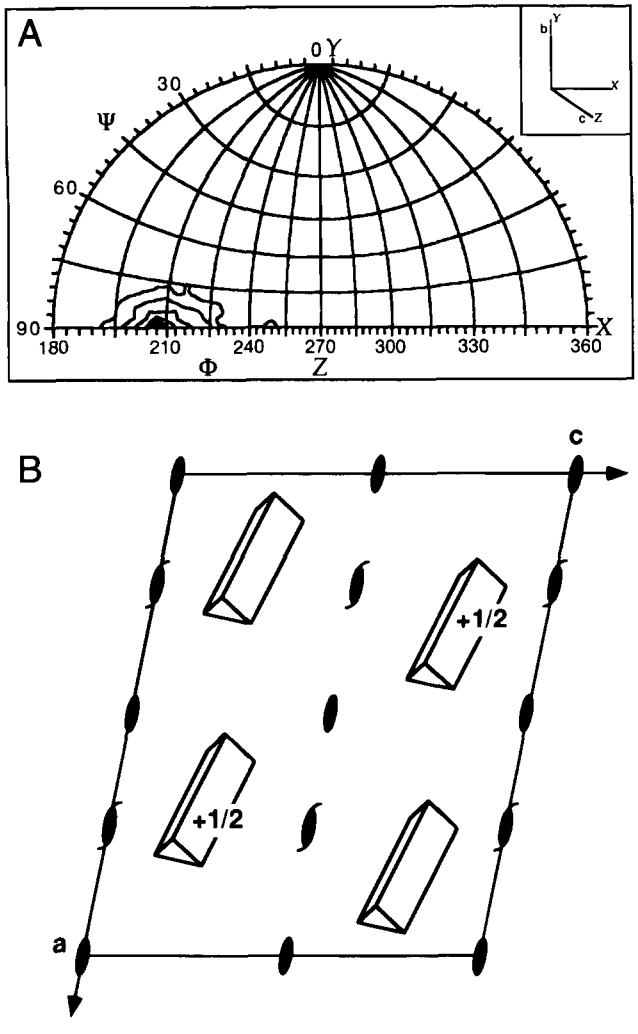

Fig. 5. A: A rotation function (Tong & Rossmann, 1990) demonstrates a noncrystallographic threefold axis that is perpendicular to the crystallographic twofold *b*-axis. A search for threefold symmetry using data from 7 to 4 Å resolution yields a major peak (11.9 σ) at $\Phi = 207.4°$ and $\Psi = 90.0°$. The same peak is found using several different fragments of the data set, varying resolution ranges and the number of large terms. The relationship between crystallographic and rotation function coordinate systems is defined in Rossmann and Blow (1962). Inset: the unit cell *b* axis (the crystallographic twofold) is coincident with the rotation function *Y* axis, the unit cell *c* axis is coincident with the rotation function *Z* axis. This self-rotation function was calculated using 1,785 reflections and 473 large terms. The contour is drawn in steps of 1.5 σ starting at 2.5 σ. **B:** The arrangement of the SIV gp41$^{27-149}$ unit cell indicating the orientation and approximate position of gp41 trimers. The trimers are represented as triangular prisms whose threefold axis is 62.5° from the *c* axis (see text). Trimer positions are dictated by C2 symmetry. Following a right-handed convention, the *b* axis goes into the plane of the figure. Twofold symmetry operators are simple ovals; operators for 2_1 screw axes are ovals with tails.

unit cell be in parallel orientations. We find no evidence for twofold and fourfold noncrystallographic symmetry in other rotation function searches. In light of the C2 symmetry and the noncrystallographic threefold axes, we tested and found that this unit cell cannot be reduced to a smaller cell with R32 symmetry (see Zlotnick et al., 1993).

Table 1. *Summary of crystallographic data*[a]

Unique reflections	Number of crystals	Total reflections	% Complete	Average I/σ	R_{merge}
2,389	2	20,644	68.2	30.3	6.6%

[a]Crystals diffract to 3.8 Å and data were collected between 20 Å and 4.0 Å. The multiple observations scale together with a R_{merge} of 6.6%, where

$$R_{merge} = 100 \times \frac{\sum_h \sum_i |I_h - I_{hi}|}{\sum_h \text{In}}$$

and I_h is the mean intensity (I) of i observations of reflection h. Because of the crystal geometry, we collected only 68.2% of the unique 3,500 reflections expected for this unit cell at 4 Å resolution.

Discussion

Protein production

Both the SIV and HIV proteins were expressed in *E. coli* at very high levels and, following protein folding, about 70–80% of the protein was recovered as soluble protein. Other investigators have commented on the low solubility of the glycosylated HIV gp41$^{21-166}$ (Weissenhorn et al., 1996) and the insolubility of the *E. coli*-produced HIV gp41$^{29-158}$ under physiological conditions (Lu et al., 1995). The formation of antibody complexes (Weissenhorn et al., 1996) or proteolytic digestion (Lu et al., 1996) were approaches used to increase or solubilize these proteins. In this report, the use of mildly acidic solvents for folding the protein was the key to both eliminating aggregation during protein folding and the ability to maintain the protein at concentrations >20 mg/mL. At solvent pH values above 3.0, based on the near-UV CD spectra, the proteins appear to have a native-like conformation and not a molten globule-like structure typical of some acid-denatured proteins (Christensen & Pain, 1994). Sedimentation equilibrium analysis and SDS-PAGE under nonreducing conditions both indicate a trimeric quaternary structure. We observed that the solubility of the protein dropped off markedly above pH 4.0; however, the solubility and conformation of the protein can be maintained in higher pH buffers by the inclusion of various alcohols, especially 2-propanol (up to 20%) and tertiary butyl alcohol (up to 50%) with no change in structure. Apart from the ability to crystallize the SIV-derived protein from acidic pH buffers, the protein at pH 3.0–3.5 is well suited for structure determination using multidimensional NMR methods (data not shown).

Subunit orientation and shape

The crystallographic data, the V_m, and the rotation function result indicate that SIV gp41 is trimeric with threefold symmetry, in agreement with the analytical ultracentifugation analyses. An important implication of the rotation function result is that the SIV gp41 trimer is a bundle of parallel subunits, related to the proposed trimer of the 43 and 53 residue fragments of gp41 reported by Blacklow et al. (1995).

We have shown that, in the wild-type SIV protein in the absence of reductant, even at pH 3.0–3.5, intermolecular disulfide bond formation occurs by air oxidation (Fig. 2, lanes G and H). This crosslinking must be between the subunits of the native trimer

rather than between subunits of different trimers, because the molecular weight of the oxidized protein was close to that of a trimer and not, for example, a hexamer (unreported analytical ultracentrifugation data). Hence, in the trimer, the α-carbon of one cysteine residue from one subunit must be in close proximity, between 4.6 and 7.4 Å (Thornton, 1981), to a cysteine from a second subunit. It follows, given that at least two intermolecular crosslinks are required to form a covalent trimer, that a second cysteine on either the first or second subunit must, in turn, be in close proximity to a cysteine on the third subunit. The disulfide crosslinking pattern supports the crystallographic data (discussed above), which indicates the subunits are orientated parallel with respect to one another, and which would cluster the cysteine containing loop regions.

If we assume that the trimeric gp41 protein is a rod-shaped molecule, then following the approach used by Beck et al. (1996; see their Fig. 2B), we can calculate predicted s values as a function of rod length. If we compare the predicted values with the actual determined s values of 3.0 for the HIV gp41$^{27-154}$, we get an estimate of about 12 nm for the length of the trimer using a hydration value of 1 g water per g protein (Beck et al., 1996). Lower hydration values result in greater predicted lengths of the rods. Electron microscopy of a glycosylated HIV gp41$^{21-166}$ produced in insect cells (Weissenhorn et al., 1996) indicated a length of 12.5 ± 1.3 nm, which is within the error of the value we have determined by hydrodynamic methods. This supports the view that the conformation of the gp41 ectodomain is not influenced markedly by glycosylation or the relatively low pH.

In trimeric coiled coil-coil structures, there is a rise of 0.145 nm per residue (Ogihara et al., 1997). Therefore, a 12-nm long coiled-coil would require about 80 residues. Kim and coworkers (Blacklow et al., 1995; Lu et al., 1995) suggested a model for the gp41 trimer based on the structure of proteolytic fragments in which each subunit is an α-helical hairpin with N-terminal 54-residue and C-terminal 46-residue helices; if correct, the length of the helical region would be about 8 nm long. In the unprocessed protein used in this study, we have only the loop regions of about 20 residues per subunit to make up the 4–4.5 nm difference in length. An alternative suggestion, based on the low-pH-induced model of the influenza virus (Bullough et al., 1994), is that each subunit in the trimer consists of one extended helix rather than a helix-turn-helix. This extended helix is estimated to contain between 90 and 100 residues. Based on the work presented herein and the analysis of crystals that diffract to 2 Å, we can expect a high-resolution structure of the gp41 ectodomain in the near future, which will resolve these issues.

Materials and methods

Plasmid construction and protein expression

DNA encoding gp41 residues Ala-27–Ser-149 from SIV$_{mac}$239 (Kestler et al., 1990) were generated as an *Nde* I–*Bam*H I fragment using PCR as described by Scharf et al. (1986) and cloned into the expression vector pET11a (Studier et al., 1990). Residues 27 and 149 correspond to 552 and 674 of the SIV *env* gene (Regier & Desrosiers, 1990). The double mutant, C86A, C92A, was made by changing the codons encoding cysteines at positions 86 and 92 to those encoding alanine using PCR (Vallette et al., 1989). DNA coding gp41 residues Thr-27–Lys-154 from HIV strain HXB2, with the cysteines at positions 87 and 93 mutated to those encoding alanine, was similarly cloned into pET11a. (Residues 27 and 154 correspond to 538 and 665 of the HIV *env* gene.) Protein expression was made with host cells BL21 (DE3) transformed with the expression vector. Fermentations in complex and minimal media were as described previously (Yamazaki et al., 1996).

Protein purification

Cells were suspended in 100 mM Tris-HCl, pH 8.0, containing 10 mM EDTA. Cell breakage and preparation of insoluble inclusion body protein were as described previously (Yamazaki et al., 1996), except that urea and Triton X-100 were excluded from the washing buffer. The final washed pellet was solubilized with 50 mM Tris-HCl, pH 8.0, containing 8 M guanidine-HCl and 50 mM DTT and the slightly cloudy solution was clarified by centrifugation at 100,000 × g for 30 min and applied at 5 mL min^{-1} to a column 6.0 cm × 60 cm of Superdex 200 (Pharmacia Biotech) equilibrated with 50 mM Tris-HCl, pH 8.0, containing 4 M guanidine-HCl and 10 mM DTT. The column was eluted at 5-mL min^{-1} and 20-mL fractions were collected. Fractions were analyzed by SDS-PAGE after removal of the guanidine-HCl using the method of Pepinsky (1991). The guanidine-HCl concentration in the protein solution was increased to about 6 M and the protein concentrated by ultrafiltration in a stirred cell with a Diaflo PM10 membrane (Millipore-Amicon) to about 20 mg/mL. Protein at this stage could be stored at −80 °C.

Protein folding

Protein, 10 mL at ~20 mg/mL in 6 M guanidine-HCl, was applied to a 1 cm × 10 cm column of Source 15 RPC (Pharmacia-Biotech) equilibrated with 0.1% TFA in water at room temperature. The protein was eluted using a linear gradient (10 column volumes) of 0–70% acetonitrile in 0.1% TFA. Wild-type protein eluted with ~40% acetonitrile, whereas the double cysteine to alanine mutant eluted with ~20% acetonitrile. The protein concentration of the eluate was adjusted to ~0.5 mg/mL with 50% acetonitrile, the solution was then dialyzed against 6 L (30–50 volumes) of 50 mM sodium formate, pH 3.0. The dialysis was performed at 4 °C and the buffer was changed at least once. The clear dialysate was concentrated as described above to 20 mg/mL and in lots of 0.5 mL applied to a column 7.5 mm × 300 mm of TSK-2000 SW (Beckman) equilibrated at room temperature with 50 mM sodium formate, pH 3.0, containing 10% 2-propanol. Protein was eluted at 3 mL/min and the second (main) peak from the column collected. The pooled protein was dialyzed against 50 mM sodium formate, pH 3.0, concentrated as described above to ~15–20 mg/mL, and sterile filtered using a Millex-GV 0.22 μm filter unit (Millipore). The folding protocol for the wild-type and mutant proteins was identical except that 2.5 mM DTT was included in the dialysis buffers for the former.

Determination of protein concentration

The protein concentration of purified proteins was determined by measuring absorbances at 280 nm in a 1-cm pathlength cell using a double-beam, diode array Hewlett-Packard 8450A UV/VIS spectrophotometer. The molar absorbance coefficients (ϵ) of native proteins were calculated from the amino acid compositions according to Wetlaufer (1962). Values of 36.8 mM^{-1} cm^{-1} ($A^{0.1\%}$ = 2.54) for both the SIV wild-type and double cysteine mutant and 42.50 mM^{-1} cm^{-1} ($A^{0.1\%}$ = 2.85) for the HIV double cysteine mutant were used.

Analytical ultracentifugation

Analytical ultracentrifugation was performed using a Beckman Optima XL-A analytical ultracentrifuge with an An-60Ti rotor and standard double-sector centerpiece cells. For equilibrium measurements, centrifugations (14–20 h at 20 °C) were at 16,000 rpm. Sedimentation velocity measurements were made at 40,000 rpm for 2–3 h at 20 °C with data collection every 15 min. Data were analyzed using both the standard Beckman XL-A data analysis software (v3.0 for DOS) and the Beckman-Origin software (v2.0 for Windows). Protein partial specific volumes were calculated from amino acid compositions (Cohn & Edsall, 1943). Values of 0.736 and 0.737 g mL^{-1} were used for the wild-type and the double cysteine mutant, respectively, and 0.738 for the HIV double cysteine mutant. Solvent densities were either calculated as described by Laue et al. (1992) or the values were taken from the International Critical Tables (1929).

CD

Spectra were recorded at 20 °C on a Jasco J-720 spectropolarimeter. Measurements in the near (340–240 nm) and far (260–180 nm) UV regions were made using 1-cm and 0.01-cm pathlength cells, respectively. A 1-nm bandwidth was used for both spectral regions. The protein solutions were about 0.7 mg mL^{-1}. Protein buffers were exchanged for 50 mM sodium formate, pH 3.0 or 3.5, or for any of the other buffers mentioned in the text, using a PD-10 column (Pharmacia Biotech). Immediately prior to use, solutions were filtered with Millex-GV 0.22-μm filter units and degassed. Secondary structures were estimated using the methods of Chang et al. (1978), Perczel et al. (1992), and Sreerama and Woody (1993).

Protein crystallization

Crystals of the SIV $gp41^{27-149}$ were grown by vapor diffusion using both the sitting and hanging drop method using Linbro plates (Flow Laboratories, McLean, Virginia) and silanized glass cover slips with and without sitting drop bridges (Hampton, Laguna Hills, California). Crystallization trials used 4 μL of a 15 mg/mL protein solution mixed with an equal volume mother liquor and 10 μL of 50 mM sodium formate, pH 3.5. Crystals, triangular and diamond-shaped plates, grew from an amorphous precipitate and appeared after about two weeks. Diffraction quality crystals were grown from 22.5% MPD, 50 mM sodium acetate, pH 4.5, 0.2 M $CaCl_2$.

X-ray data collection and processing

Crystals grown in 22.5% MPD, 50 mM sodium acetate, pH 4.5, 0.2 M $CaCl_2$ were flash frozen with liquid propane, using their mother liquor as a cryo-protectant. Crystals were maintained at −140 °C during data collection with a Molecular Structure Corp. crystal cooling system. Data were collected with an R-axis IIC image plate detector mounted on a Rigaku RU2000 X-ray generator equipped with focusing mirrors. Data were collected in 2° oscillations and 20-min exposures. Diffraction data were autoindexed and processed using the program DENZO (Otwinowski, 1993). Data from two crystals were scaled and merged together using Scalepack (Table 1). Rotation functions were calculated with intensity data using the program GLRF (Tong & Rossmann, 1990).

Acknowledgments

We thank Ira Palmer for assistance with protein purification and Pat Spinella for the N-terminal amino acid sequencing. This work was supported by the AIDS Targeted Antiviral Program of the Office of the Director of the National Institutes of Health (S.J.S. and P.T.W.).

References

Beck K, Gambee JE, Bohan CA, Bachinger HP. 1996. The C-terminal domain of cartilage matrix protein assembles into a triple-stranded α-helical coiled-coil structure. *J Mol Biol 256*:909–923.

Blacklow SC, Lu M, Kim PS. 1995. A trimeric subdomain of the simian immunodeficiency virus envelope glycoprotein. *Biochemistry 34*:14955–14962.

Bosch ML, Earl PL, Fargnoli K, Picciafuoco S, Giombini F, Wong-Staal F, Franchini G. 1989. Identification of the fusion peptide of primate immunodeficiency viruses. *Science 244*:694–697.

Bour S, Geleziunas R, Wainberg MA. 1995. The human immunodeficiency virus type 1 (HIV-1) CD4 receptor and its central role in promotion of HIV-1 infection. *Microbiol Rev 59*:63–93.

Bullough PA, Hughson, FM, Skehel JJ, Wiley DC. 1994. Structure of influenza haemagglutin at the pH of membrane fusion. *Nature 371*:37–43.

Capon DJ, Ward RH. 1991. The CD4-gp120 interaction and AIDS pathogenesis. *Annu Rev Immunol 9*:649–678.

Chang CT, Wu CSC, Yang JT. 1978. Circular dichroism analysis of protein conformation: Inclusion of the β-turns. *Anal Biochem 91*:13–31.

Christensen R, Pain RH. 1994. The contribution of the molten globule model. In: Pain RH, ed. *Mechanisms of protein folding*. New York: IRL Press at Oxford University Press Inc. pp 55–79.

Cohn EJ, Edsall JT. 1943. *Proteins, amino acids and peptides*. Princeton, New Jersey: Van Nostrand-Reinhold. p 370.

Geleziunas R, Bour S, Wainberg MA. 1994. Human immunodeficiency virus type 1-associated CD4 downregulation. *Adv Virus Res 44*:203–266.

Higashi T. 1990. *Process: A program for indexing and processing R-axis II image plate data*. Tokyo: Rigaku Corp.

Kestler HW, Kodama T, Ringler D, Marthas M, Pedersen N, Ratner A, Regier D, Sehgal, T, Daniel M, King N, Desrosiers RC. 1990. Induction of AIDS by molecularly cloned virus. *Science 248*:1109–1112.

Kowalski M, Bergeron L, Dorfman T, Haseltine W, Sodroski J. 1991. Attenuation of human immunodeficiency virus type 1 cytopathic effect by a mutation affecting the transmembrane envelope glycoprotein. *J Virol 65*:281–291.

Laue TM, Shah B, Ridgeway TM, Pelletier SM. 1992. Computer-aided intrepretation of sedimentation data for proteins. In: Harding SE, Rowe AJ, Horton JC, eds. *Analytical ultracentrifugation in biochemistry and polymer sciences*. Cambridge, UK: Royal Society of Chemistry.

Lu M, Blacklow SC, Kim PS. 1995. A trimeric structural domain of the HIV-1 transmembrane glycoprotein. *Nature Struct Biol 2*:1075–1082.

Matthews BW. 1968. Solvent content of protein crystals. *J Mol Biol 33*:491–497.

Monera OD, Zhou NE, Kay CM, Hodges RS. 1993. Comparison of antiparallel and parallel two-stranded helical coil-coils. *J Biol Chem 268*:19218–19227.

Ogihara NL, Weiss MS, DeGrado WF, Eisenberg D. 1997. The crystal structure of the designed trimeric coiled coil-coil-V_aL_d: Implications for engineering crystals and supermolecular assemblies. *Protein Sci 6*:80–88.

Otwinowski Z. 1993. Oscillation data reduction program. In: Sawyer L, Isaacs N, Bailey S, eds. *Proceedings of the CCP4 study weekend: Data collection and processing*. Warrington, UK: SERC Daresbury Laboratory. pp 56–62.

Pepinsky RB. 1991. Selective precipitation of proteins from guanidine hydrochloride-containing solutions with ethanol. *Anal Biochem 195*:177–181.

Perczel A, Park K, Fasman GD. 1992. Analysis of the circular dichroism spectrum of proteins using the convex constraint algorithm: A practical guide. *Anal Biochem 203*:83–93.

Regier DA, Desrosiers RC. 1990. The complete nucleotide sequence of a pathogenic molecule clone of simian immunodeficiency virus. *AIDS Res Hum Retroviruses 6*:1221–1231.

Rossmann MG. 1990. The molecular replacement method. *Acta Crystallogr A 46*:73–82.

Rossmann MG, Blow DM. 1962. The detection of sub-units within the crystallographic asymmetric unit. *Acta Crystallogr 15*:24–31.

Scharf SJ, Horn GT, Erlich HA. 1986. Direct cloning and sequence analysis of enzymically amplified genomic sequences. *Science 233*:1076–1078.

Sreerama N, Woody RW. 1993. A self-consistent method for the analysis of protein secondary structure from circular dichroism. *Anal Biochem 209*: 32–44.

Studier FW, Rosenberg AH, Dunn JJ. 1990. Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol 185*:60–89.

Thornton JM. 1981. Disulphide bridges in globular proteins. *J Mol Biol 151*:261–287.

Tong L, Rossmann MG. 1990. The locked rotation function. *Acta Crystallogr A 46*:783–792.

Vallette F, Merge E, Reiss A, Adesnik M. 1989. Construction of mutant and chimeric genes using the polymerase chain reaction. *Nucleic Acid Res 17*:723–733.

Van Holde KE. 1975. Sedimentation analysis of proteins. In: Neurath R, Hill R. eds. *The proteins, 3rd ed, vol 1*. New York: Academic Press. pp 225–291.

Washburn EW, ed. 1929. *International critical tables of numeric data, physics, chemistry and technology*. New York and London: McGraw-Hill Book Company Inc.

Weissenhorn W, Wharton SA, Calder LJ, Earl PL, Moss B, Aliprandis E, Skehel JJ, Wiley DC. 1996. The ectodomain of HIV-1 env subunit gp41 forms a soluble, α-helical, rod-like oligomer in the absence of gp120 and the N-terminal fusion peptide. *EMBO J 15*:1507–151.

Wetlaufer DB. 1962. Ultraviolet spectra of proteins and amino acids. *Adv Prot Chem 17*:303–390.

Yamazaki T, Hinck AP, Wang YX, Nicholson LK, Torchia DA, Wingfield PT, Stahl SJ, Kaufman JD, Chang CH, Domaille PJ, Lam PYS. 1996. Three-dimensional solution structure of the HIV-1 protease complexed with DMP323, a novel cyclic urea-type inhibitor, determined by nuclear magnetic resonance spectroscopy. *Protein Sci 5*:495–506.

Zlotnick A, McKinney BR, Munshi S, Bibler J, Rossmann MG, Johnson JE. 1993. A monoclinic crystal with R32 pseudo-symmetry: A preliminary report of Nodamura virus structure determination. *Acta Crystallogr D 49*:580–587.