Crystallization of a trimeric human T cell leukemia virus type 1 gp21 ectodomain fragment as a chimera with maltose-binding protein

ROB J. CENTER,¹ BOSTJAN KOBE,¹ KIRILEE A. WILSON,¹ TRAZEL TEH,¹
GEOFFREY J. HOWLETT,² BRUCE E. KEMP,¹ AND PANTELIS POUMBOURIOS¹

¹St. Vincent's Institute of Medical Research, 41 Victoria Parade, Fitzroy 3065, Australia

²Department of Biochemistry and Molecular Biology, University of Melbourne, Parkville 3052, Australia

(RECEIVED February 3, 1998; ACCEPTED March 16, 1998)

Abstract

We present a novel protein crystallization strategy, applied to the crystallization of human T cell leukemia virus type 1 (HTLV-1) transmembrane protein gp21 lacking the fusion peptide and the transmembrane domain, as a chimera with the *Escherichia coli* maltose binding protein (MBP). Crystals could not be obtained with a MBP/gp21 fusion protein in which fusion partners were separated by a flexible linker, but were obtained after connecting the MBP C-terminal α -helix to the predicted N-terminal α -helical sequence of gp21 via three alanine residues. The gp21 sequences conferred a trimeric structure to the soluble fusion proteins as assessed by sedimentation equilibrium and X-ray diffraction, consistent with the trimeric structures of other retroviral transmembrane proteins. The envelope protein precursor, gp62, is likewise trimeric when expressed in mammalian cells. Our results suggest that MBP may have a general application for the crystallization of proteins containing N-terminal α -helical sequences.

Keywords: *E. coli* protein expression; maltose-binding protein/HTLV-1 gp21 chimera; protein crystallization; trimerization; X-ray diffraction

Crystallization and phase determination remain the two rate-limiting steps in determining protein crystal structures. The strategy of using fusion proteins comprising a "crystallization tag" of known three-dimensional structure and the protein of interest offers an approach to overcoming these problems. The crystallization conditions and crystal contacts found important in crystallization of the tag may be used to guide the crystallization of the fusion protein. Knowledge of the three-dimensional structure of the crystallization tag should then allow determination of the starting phases by the method of molecular replacement. Moreover, knowledge of the heavy atom-binding sites in the crystallization tag can be used to determine the starting phases by the methods of multiple isomorphous replacement and multi-wavelength anomalous dispersion. The strategy offers similar advantages as co-crystallization with antibody fragments (Air et al., 1987; Prongay et al., 1990; Jacobo-Molina et al., 1991; Ostermeier et al., 1995) while being

less laborious. However, reports of successful crystallization and structure determination of fusion proteins have thus far been limited to small peptides fused to much larger fusion tags (e.g., Donahue et al., 1994; Lim et al., 1994). Most commonly used fusion protein expression systems contain a linker sequence between the fusion protein components. The presence of the flexible linker and the lack of complementary surfaces may prevent stable association of the fusion partners. We reasoned that by eliminating the linker and engineering a rigid connection between the fusion partners we may facilitate crystallization.

The native HTLV-1 envelope protein (Env) is synthesized as a polyprotein precursor, gp62, that is cleaved by a cellular protease to produce the receptor-binding subunit, gp46 (SU), and the transmembrane protein (TM) subunit, gp21, which remain associated noncovalently (Pique et al., 1992; Paine et al., 1994). The SU-TM oligomeric complex is incorporated into budding virions at the plasma membrane. Unlike other retroviruses, cell-free HTLV-1 virions have very low infectivity (Fan et al., 1992); viral spread is, therefore, likely to occur via Env-mediated fusion between infected cells and target cells (Donegan et al., 1994; Delamarre et al., 1997). Receptor binding is presumed to induce conformational changes in the SU-TM complex that result in TM-mediated fusion between the infected cell and target cell membranes (reviewed in Delamarre et al., 1996). The cellular receptors required for HTLV-1 fusion are not yet known.

Reprint requests to: Pantelis Pountourios, St. Vincent's Institute of Medical Research, 41 Victoria Parade, Fitzroy 3065, Australia; e-mail: apoum@ariel.its.unimelb.edu.au.

Abbreviations: Env, envelope protein; GST, glutathione-S transferase; HA, hemagglutinin; HTLV-1, human T cell leukemia virus type 1; MAbs, monoclonal antibody; MBP, maltose binding protein; MoMuLV, Moloney murine leukemia virus; SU, receptor-binding subunit; TM, transmembrane subunit.

Retroviral TMs have a similar structural organization despite limited sequence similarities (Gallagher et al., 1989). Conserved elements include an N-terminal hydrophobic fusion peptide followed by a putative coiled coil-forming sequence, a disulfide-bonded loop, a region with one or several N-linked glycosylation sites, and a hydrophobic membrane-spanning sequence that anchors the SU-TM complex to membranes (Fig. 1A). Retroviral TMs also share structural features with the influenza virus hemagglutinin (HA) TM (HA2), for which three-dimensional structures of pre-fusogenic and fusion-activated forms are known. Both structures are assembled into trimers via a triple-stranded coiled coil within the HA core (Wilson et al., 1981; Bullough et al., 1994). Fusion activation, triggered by exposure of HA to low pH in the endosome causes displacement of the receptor-binding sub-

unit (HA1), allowing the transition of HA2 from the metastable, pre-fusogenic conformation to the most energetically favorable, fusion-activated structure (Bullough et al., 1994). This involves helical extension of the triple-stranded coiled coil by sequences previously in an extended conformation, thereby relocating the N-terminal fusion peptide by over 100 Å. Structural analyses of intact retroviral TMs have been hampered by their low solubility when expressed in *Escherichia coli* (Fass et al., 1996; Chan et al., 1997; Weissenhorn et al., 1997b). However, crystallizable fragments of Moloney murine leukemia virus (MoMuLV) (Fass et al., 1996) and HIV-1 (Chan et al., 1997; Tan et al., 1997; Weissenhorn et al., 1997b) TMs were identified by protease dissection after their expression in the absence of SU. Determination of the structures of recombinant and synthetic peptides derived from these studies revealed trimeric coiled-coil cores analogous to fusion-activated HA2. These fragments, however, lack either a substantial C-terminal portion (Fass et al., 1996) or a region containing the conserved disulfide-bonded loop (Chan et al., 1997; Tan et al., 1997; Weissenhorn et al., 1997b) shown to be critical for the infectivity of several retroviruses (Syu et al., 1991; Pancino et al., 1995; Thomas & Roth, 1995).

Here we report the high-yield expression in *E. coli* of soluble fusion proteins comprising MBP and continuous segments of the gp21 ectodomain. Crystallization of MBP/gp21 chimeras was achieved following fusion of the MBP C-terminal α -helix to the predicted gp21 coiled-coil-forming sequence via three alanine residues. The X-ray diffraction data together with sedimentation equilibrium and chemical crosslinking of the Env precursor expressed in COS-7 cells revealed the trimeric quaternary structure of HTLV-1 Env.

Results

The MBP/gp21(335–445) chimera acquires trimeric structure

Expression in *E. coli* of MBP/gp21(335–445) comprising MBP fused to the gp21 ectodomain fragment, Thr-335–Thr-445, which lacks the N-terminal fusion peptide and C-terminal transmembrane sequence (Fig. 1B), resulted in high yields of soluble protein. The oligomerization state of MBP/gp21(335–445) eluted from amylose-agarose was examined by Superdex 200 gel filtration (Fig. 2A, top panel). The majority of MBP/gp21(335–445) eluted as a symmetrical peak (elution time of 58 min) indicating the acquisition of a discrete oligomeric structure. A low level of high-molecular weight aggregate (co-eluting with blue dextran) was also observed. Maltose-binding protein expressed in the absence of gp21 had an elution time of 81 min (Fig. 2A, middle panel), similar to that of the 43 kDa marker, ovalbumin (77 min) and is consistent with monomeric structure (Spurlino et al., 1991). The gp21 domain, therefore, confers oligomeric structure to MBP/gp21(335–445). Fractions corresponding to MBP/gp21(335–445) oligomers from multiple gel filtration runs were pooled, concentrated, and re-chromatographed (Fig. 2A). SDS-PAGE analysis of the purified MBP/gp21(335–445) oligomer in the presence of β -mercaptoethanol indicated a single band migrating at ~53 kDa (Fig. 2B), whereas SDS-PAGE in the absence of β -mercaptoethanol showed a low level of disulfide-linked dimers. To accurately determine the quaternary structure of MBP/gp21(335–445), the peak fraction from a gel filtration experiment was analyzed by sedimentation equilibrium. The observed native molecular weight of MBP/gp21(335–445) oligomer was 167 kDa (data not shown), which corresponds

Fig. 1. **A:** Schematic representation of the gp21 sequence showing the N-terminal fusion peptide, the major predicted α -helices, the disulfide-bonded loop (C CC), the single N-linked glycosylation site (Ψ), and the transmembrane sequence (TM). Schematic diagrams of the MBP/gp21 fusion proteins examined are shown below. The numbers in brackets refer to the N- and C-terminal residues of the gp21 sequences included in each construct. Sequences used to link the MBP and gp21 moieties are S₃N₁₀LGIEGRISFEFGS (Linker) in MBP/gp21(335–445) or three alanine residues (AAA) in MBPL-/gp21(338–445) and MBPL-/gp21(338–425). **B:** The amino acid sequence of MBP/gp21(335–445) showing the MBP C-terminal α -helix, the linker sequence, and gp21 residues Thr-335–Thr-445. Grey underlining represents predicted α -helices (program PHD; Rost, 1996). **C:** MBP and gp21 linked via three alanine residues in the constructs MBPL-/gp21(338–445) and MBPL-/gp21(338–425). Amino acid changes introduced into the MBP C-terminal α -helix are underlined.

Fig. 2. A: Gel filtration profiles of MBP/gp21(335–445) (top panel) and MBP (middle panel) after elution from amylose agarose, and MBP/gp21(335–445) oligomers after purification by gel filtration chromatography (bottom panel). The elution position of the calibration markers ferritin (440 kDa), catalase (232 kDa), and ovalbumin (43 kDa) are shown by triangles. The void volume (V_0) was determined using blue dextran. B: SDS-PAGE analysis of MBP/gp21(335–445). Lanes 1 and 3: eluate from amylose-agarose; lanes 2 and 4: oligomer purified by gel filtration chromatography. Proteins were subjected to SDS-PAGE on 10% polyacrylamide gels in the presence (lanes 1 and 2) or absence (lanes 3 and 4) of 1% β -mercaptoethanol. Gels were stained with Coomassie brilliant blue.

closely to a complex of three MBP/gp21(335–445) molecules (166.8 kDa).

Retroviral Env oligomerization is mediated by the TM protein (Einfeld & Hunter, 1994). To determine if the oligomerization state of MBP/gp21(335–445) is consistent with the quaternary structure of the HTLV-1 Env precursor, gp62, COS-7 cell lysates containing gp62 were subjected to sucrose density gradient centrifugation to separate oligomeric species. Figure 3 (upper panel) indicates that gp62 sediments in a broad peak corresponding to a range of oligomeric structures, as observed previously for Env precursors of HTLV-1 and other retroviruses (Paine et al., 1991; Einfeld & Hunter, 1994; Center et al., 1997). Chemical crosslinking of gp62 oligomers by treating the gradient fractions with BS³ resulted in bands corresponding to dimer and trimer in the peak fractions (Fig. 3, lower panel). The dimer band migrated as a doublet that may have arisen from the migration of crosslinked dimers of different shapes in SDS-PAGE experiments. BS³-treated gp62 sedimenting faster than the 232 kDa marker did not enter the separating gel, suggesting it corresponds to large gp62 aggregates. These results indicate that the highest order quaternary structure for gp62 is a trimer, consistent with that observed for MBP/gp21(335–445).

Fig. 3. Sucrose density gradient sedimentation and crosslinking profiles of gp62. COS-7 cells were transfected with wild-type Env encoding plasmid then infected with vaccinia virus WR and lysed 24 h post-transfection. Cell lysates were subjected to centrifugation through 5–20% sucrose gradients. Gradients were fractionated and proteins crosslinked with 1.0 mM BS³ where indicated. Trichloroacetic acid-precipitated proteins were separated by SDS-PAGE on 5–15% gradient gels under reducing conditions. Proteins were transferred to nitrocellulose, immunoblotted with MAb C8, and detected by addition of ¹²⁵I-protein A and PhosphorImager analysis. gp62 species are indicated by arrowheads. The direction of sedimentation is right to left, and is indicated by an arrow. Unfractionated cell lysates (CL) were also subjected to SDS-PAGE to mark the position of monomeric gp62.

N- and C-terminal modification of MBP/gp21(335–445)

Extensive crystallization trials did not yield crystals of MBP/gp21(335–445). To determine if this was due to the linker peptide separating MBP and gp21 sequences (Fig. 1B) introducing flexibility to the fusion protein, and thus inhibiting crystal formation, we fused the C-terminal α -helix of MBP to the predicted N-terminal α -helix of gp21 via three alanine residues in the construct MBPL-/gp21(338–445) (Fig. 1C). Alanines were chosen as the linking amino acids due to their helix-stabilizing properties (Heinz et al., 1992; Blaber et al., 1994). Three charged amino acids in the C-terminal α -helix of MBP (Glu-360, Lys-363, and Asp-364) were also replaced by alanines to avoid potential electrostatic repulsion (based on molecular modeling; program "O"; Jones et al., 1990) between MBP monomers when brought into close proximity by the trimerization of the HTLV-1 gp21 moieties. We also produced the C-terminally truncated construct MBPL-/gp21(338–425), as secondary structure predictions for gp21 (Fig. 1B) suggested a region with high helical propensity terminated at Thr-425. Further modifications were not made, as C-terminal truncation to Asn-407 resulted in aggregation of the fusion protein (data not shown).

Oligomers of MBPL-/gp21(338–445) and MBPL-/gp21(338–425) were purified to homogeneity as assessed by Superdex 200 gel filtration (data not shown). The trimeric structure of MBPL-/gp21(338–425) was confirmed by sedimentation equilibrium (Fig. 4). The native molecular weight was calculated from global fitting of data obtained using starting concentrations of 0.5, 0.167, and 0.056 mg/mL. Figure 4 shows the fitted data and residuals calculated for a molecular weight value of 153 kDa, closely approximating that expected for a trimer (151.2 kDa).

Redox state of the MBP/gp21(338–425) chimera

HTLV-1 gp21 contains three conserved ectodomain cysteines at positions 393, 400, and 401. To investigate the redox state of MBP/gp21(338–425) cysteines, their susceptibility to alkylation by 4-vinylpyridine (105.1 Da) was assessed by electrospray mass spectrometry. The mass of MBP/gp21(338–425) was determined as $50,432 \pm 2$ Da (Fig. 5A), which closely approximates its theoretical molecular weight, 50,431.7 Da. Alkylation of MBP/gp21(338–425) led to a mass increase of 106 Da ($50,538 \pm 2$ Da) (Fig. 5B), consistent with the presence of one unpaired cysteine. Prior reduction of MBP/gp21(338–425) with 10 mM dithiothreitol followed by alkylation with 50 mM 4-vinylpyridine resulted in the presence of two species. The major species ($50,747 \pm 2$ Da) is consistent with the alkylation of three cysteines, while the minor species ($50,647 \pm 2$ Da) has two modified cysteines resulting from incomplete alkylation (Fig. 5C). These results indicate the presence of a disulfide bond in the MBP/gp21(338–425) chimera, which by analogy with the TM of MoMuLV (Fass et al., 1996), may be formed through the pairing of Cys 393 and Cys 400.

Crystallization and crystallographic data collection

Crystals were obtained for the constructs MBPL-/gp21(338–445) and MBPL-/gp21(338–425) in which fusion partners were linked by three alanines. MBPL-/gp21(338–445) crystallized in two forms: plates of $0.5 \times 0.5 \times 0.01$ mm³ in 20% PEG 10,000/0.1 M HEPES, pH 7.5 at 4°C and room temperature; and thin needles in 18% PEG 8,000/0.2 M zinc acetate/0.1 M sodium cacodylate, pH 6.5 at 4°C. MBPL-/gp21(338–425) yielded cuboidal crystals

Fig. 4. Sedimentation equilibrium profiles for MBPL-/gp21(338–425). The solutions in S-buffer were centrifuged at 12,000 rpm and 20°C for 16 h. The optical density at 280 nm as a function of radial distance was fitted by global analysis assuming a single solute. The residuals of the fitted lines to the experimental data are given in A–C. The experimental data and the fitted lines in D refer to starting concentrations of 0.5 (open circles), 0.17 (open squares), and 0.055 (open triangles) mg/mL.

of $0.3 \times 0.3 \times 0.2$ mm³ in 22% PEG 4,000/0.2 M ammonium sulfate/0.1 M sodium acetate, pH 4.7–4.8, by mixing 2 μ L of protein (18 mg/mL) with 2.5–3.0 μ L of reservoir solution.

The cuboidal crystals of MBPL-/gp21(338–425) were suitable for diffraction data collection; they have the symmetry of the rhombohedral space group R3 with cell dimensions $a = b = 102.0$ Å, $c = 118.2$ Å, and angles $\alpha = \beta = 90^\circ$, $\gamma = 120^\circ$. 262,939 diffraction measurements were collected at DESY synchrotron from two crystals to yield 11,930 unique reflections between infinity and 2.44 Å resolution; the data set was 92% complete with $R_{\text{merge}} = 0.145$ between infinity and 3 Å resolution ($R_{\text{merge}} = 0.065$ between infinity and 6.46 Å resolution). Crystals most likely contain one monomer per asymmetric unit; the corresponding Matthews coefficient (2.34 Å³/Da) and solvent content (47%) are within the range observed for protein crystals (Matthews, 1968). The crystal symmetry is consistent with the trimeric quaternary structure (Fig. 6).

Discussion

Continuous segments of the gp21 ectodomain, lacking the hydrophobic fusion peptide and transmembrane sequence, formed trimers when expressed in *E. coli* as MBP fusion proteins. Crystals suitable for three-dimensional structure determination were obtained for the gp21 fragment, Met-338–Thr-425, which represents

Fig. 5. Electrospray mass spectrometry of MBP/gp21(338–425) (A), alkylated MBP/gp21(338–425) (B), and reduced and alkylated MBP/gp21(338–425) (C). The data were collected on a PE Sciex API III mass spectrometer. MBP/gp21(338–425) was alkylated with 4-vinylpyridine or reduced with dithiothreitol prior to alkylation, as described in Materials and methods. Protein was prepared for mass spectrometry by precipitation in methanol/chloroform followed by direct infusion into the mass spectrometer in 15% (v/v) acetic acid, 50% (v/v) acetonitrile.

over 80% of the ectodomain, when the MBP C-terminal α -helix was connected to the gp21 N-terminal α -helical sequence via three alanine residues. By contrast, MBP/gp21(335–445) in which fusion partners are separated by the linker sequence, Ser₃-Asn₁₀-Leu-Gly-Ile-Glu-Gly-Arg-Ile-Ser-Glu-Phe-Gly-Ser, did not crystallize. Removal of the linker presumably results in a rigid connection between the MBP and gp21 moieties, limiting the relative mobility of the two domains, thereby allowing crystal growth. Knowledge of the MBP three-dimensional structure is now being used to determine the starting phases by the method of molecular replacement. Our findings provide a strategy for the crystallization of other proteins with an N-terminal α -helical sequence as MBP fusion proteins.

Crystallization tags of known tertiary structure have been used for the crystallization of short peptides and subsequent determination of their structures (Carter et al., 1994; Donahue et al., 1994; Lim et al., 1994). However, use of fusion tags for the crystallization of substantial protein domains is limited to the 100-residue DNA-binding domain of DNA replication-related element-binding factor fused to glutathione-S-transferase (GST) (Kuge et al., 1997). The GST and DNA replication-related element-binding factor moieties were linked by the sequence Asp-Leu-Val-Pro-Arg-Gly-Ser. Because the structure of this linker is not known, the general applicability of the GST tag to carrier protein-driven crystallization is difficult to predict. The use of MBP as a crystallization tag is, therefore, advantageous because the presence of the MBP C-terminal helix allows rigid linkages between fusion protein components to be designed. Moreover, the MBP moiety may confer solubility to its fusion partner. We observed that gp21 and HIV and feline immunodeficiency virus TMs expressed as MBP fusion proteins were soluble, whereas analogous GST fusion proteins were insoluble (data not shown). Furthermore, other investigators have shown that expression of retroviral TMs without fusion partners in *E. coli* results in insoluble aggregates requiring limited proteolysis (Blacklow et al., 1995; Fass & Kim, 1995; Lu et al., 1995; Weiss-

senhorn et al., 1997a) or refolding from denaturants (Wingfield et al., 1997) to identify soluble TM fragments suitable for crystallization.

The MBP/gp21 fusion proteins described here have a trimeric quaternary structure, consistent with that observed for Env precursor, gp62, expressed in COS-7 cells. As MBP expressed in the absence of gp21 is monomeric, sequences within gp21 confer oligomerization, and the presence of MBP does not interfere with the adoption of a native quaternary structure. These results indicate that the monomeric nature of MBP (Spurlino et al., 1991) makes it a suitable crystallization carrier for oligomeric proteins. By contrast, GST is dimeric (Lim et al., 1994) and fusion with an oligomeric protein may result in a chimera with a non-native quaternary structure. Oligomerization of MBP/gp21 chimeras is most likely driven by the formation of a trimeric coiled coil by the N-terminal α -helical sequence of each monomer, analogous MoMuLV, and HIV Env (Fass et al., 1996; Center et al., 1997; Chan et al., 1997; Pombourios et al., 1997; Weissenhorn et al., 1997b). In the case of influenza virus, the trimeric coiled coil forms the core of the TM structure in both the pre-fusogenic and fusion-activated forms (Wilson et al., 1981; Bullough et al., 1994).

MBPL-/gp21(338–425) contains the putative N-terminal triple-stranded coiled coil, the disulfide-bonded domain, and a C-terminal helical element, and represents the largest continuous retroviral TM fragment crystallized to date. The MoMuLV TM structure contains an intact disulfide-bonded loop (Fass et al., 1996) but lacks predicted C-terminal α -helical elements and, consequently, does not locate the loop in an intact structural context. In the case of HIV-1 gp41, proteolysis was necessary to yield a soluble product, removing the region containing the disulfide-bonded loop, as well as amino acids at the C-terminus (Chan et al., 1997; Tan et al., 1997; Weissenhorn et al., 1997b). Alternatively, cysteine residues of the disulfide-bonded loop were mutated to alanine to support crystallization of the simian immunodeficiency virus TM (Wingfield et al., 1997). As MBPL-/gp21(338–425) contains all of the

Fig. 6. A pseudo-precession image of the *hk0* zone of the MBPL-/gp21(338–425) crystals (generated by HKLVIEW; CCP4, 1994). The image shows threefold crystallographic symmetry consistent with a trimeric quaternary structure of HTLV-1 gp21.

above elements, we anticipate that the three-dimensional structure will extend our understanding of retroviral TM structure and function and suggest avenues for the design of inhibitors of viral infection.

Materials and methods

Expression vector construction

A gene fragment encoding gp21 amino acids Thr-335–Thr-445, flanked by 5′-*Bam*HI and 3′-*Pst*I restriction sites was PCR amplified using the primer pair 5′GGCTGGCGGATCCACCGGCTCCATGTCCC and 5′CAACAAGCTGCAGTTAAGTTTGTAAGGCCTCTCG (restriction sites in bold), and pDT1 DNA containing the HTLV-1 *env* gene as the template. The PCR product was ligated into the *Bam*HI/*Pst*I sites of the MBP-expression vector pMAL-c2 (New England Biolabs, Beverly, Massachusetts) to produce pMBP/gp21(335–445) (Fig. 1B). This vector was modified for expression of other MBP/gp21 fusion proteins in which the

MBP C-terminus was fused to Met-338 of gp21 via three alanines (Fig. 1C). A gene fragment encoding a mutated version of MBP flanked by 5′*Bgl*II and 3′*Pvu*II restriction sites was PCR amplified using the primer pair 5′CTGATTTATAACAAAGATCTGCTGCCG and 5′GTTGTTTTACGCAGCTGCATTAGTCTGCGCGGCTGCCAGGGCTGCATCGACAGTC and pMAL-c2 as the template. In this fragment, the *Sac*I site in pMAL-c2, which adds three serine residues to the C-terminus of MBP, was replaced with *Pvu*II, which encodes three alanines to favor helix extension. The PCR primer also encodes the additional MBP substitutions, Glu-359 to Ala (E359A), K362A and D363A (underlined). A gp21 gene fragment encoding the amino acid sequence Met-338–Thr-445, and flanked by 5′*Pvu*II and 3′*Pst*I restriction sites, was PCR amplified with the oligonucleotide primer pair 5′GGGATTGCAGCTGCCGATGTCCTCGCCTCAGGAAAG and 5′CAACAAGCTGCAGTTAAGTTTGTAAGGCCTCTCG. The two PCR products were then ligated simultaneously into the *Bgl*II-*Pst*I sites of pMAL-C2 to produce pMBPL-/gp21(338–445). A gene fragment encoding MBPL-/gp21(338–425) was produced as for MBPL-/gp21(338–

445) except that the primer pair 5'GGGATTGCAGCTGCGATGTCCCTCGCCTCAGGAAAG and 5'CCAGTTAAGCTGCAGTTAAGTCAGGACTCG were used to PCR amplify the gp21 gene fragment. The sequences of PCR-amplified gp21 gene fragments were confirmed by the dideoxy chain termination sequencing procedure. To facilitate detection of HTLV-1 gp62 expressed in COS-7 cells, a modified version of the monoclonal antibody (MAb) C8 epitope, Pro-Arg-Gly-Pro-Asp-Arg-Pro-Glu-Gly-Ile-Glu-Glu-Gly-Gly-Glu-Arg-Asp-Arg-Asp-Arg-Ser-Ile-Arg-Leu-Val-Asn-Gly-Ser (Abacioglu et al., 1994), was fused to the C-terminal Leu-488 of gp62 through an engineered *SacI* site. The tagged *env* gene was ligated into the *SalI-EagI* sites of vector pPE16 for transient expression from the vaccinia virus early-late promoter, P7.5 (Earl et al., 1990).

Fusion protein expression and purification

Fusion proteins were expressed in *E. coli* strain BL21 (DE3/pLysS) cells. Bacteria were lysed by sonication in S-buffer (300 mM NaCl, 100 mM Tris pH 8.0, 1 mM EDTA, 0.02% sodium azide) and cell debris pelleted by centrifugation at $39,000 \times g$ for 30 min. MBP/gp21 fusion proteins were affinity purified from the soluble fraction using amylose-agarose (New England Biolabs) as recommended by the manufacturer; elution was achieved with 10 mM maltose. MBP/gp21 fusion protein oligomers were then isolated by Superdex 200 gel filtration chromatography (Pharmacia Biotech, Uppsala, Sweden) in S-buffer. Protein purity was assessed by SDS/PAGE.

Sedimentation equilibrium

Purified MBP-gp21 oligomers were subjected to sedimentation equilibrium using starting concentrations in the range 0.5–1.0 mg/mL, on a Beckman XL-A analytical ultracentrifuge equipped with a Ti60 rotor and filled-epon centerpieces (12 mm path length). MBP/gp21(335–445) and MBPL-(338–425) were centrifuged at 20°C for 16 h, at 10,000 and 12,000 rpm, respectively; the final equilibrium distributions were determined from absorption measurements at 280 nm. Theoretical partial specific volumes calculated from the amino acid composition of fusion proteins were: 0.740 for MBP/gp21(335–445) and 0.743 for MBPL-/gp21(338–425); the solvent density of S-buffer was calculated to be 1.012 g mL^{-1} . To calculate the molecular weight, data were fitted to a single ideal species model using the SEDEQ1B program (generously provided by Allen Minton, NIH, Bethesda, Maryland).

Chemical crosslinking of HTLV-1 gp62 expressed in COS-7 cells

COS-7 cells (250,000 cells per 9 cm^2 well) were transfected with plasmid DNA using lipofectamine (Life Technologies, Gaithersburg, Maryland) according to the manufacturer's instructions, and infected with the wild-type vaccinia virus WR. Cells were lysed 24 h post-transfection with PBS/1% Triton X-100, clarified by centrifugation at $10,000 \times g$ for 10 min and sedimented in 5–20% sucrose gradients containing PBS/1% Triton X-100 in a Beckman SW41Ti rotor at 35,000 rpm for 19 h at 4°C. Proteins in 0.5 mL fractions were crosslinked with 1.0 mM BS³ (Pierce Chemical Company, Rockford, Illinois) for 1 h on ice, quenched by 100 mM glycine pH 7.8 for 1 h on ice, precipitated with trichloroacetic acid, and separated by electrophoresis on 5–15% polyacrylamide gradi-

ent gels in the presence of SDS under reducing conditions. Proteins were transferred to nitrocellulose, immunoblotted with MAb C8, and probed with rabbit anti-mouse immunoglobulins (Dako A/S, Glostrup, Denmark) followed by radioiodinated protein A. Immunoblots were visualized by scanning in a PhosphorImager SF (Molecular Dynamics, Sunnyvale, California). Sucrose gradients were calibrated with the molecular weight markers catalase (11.3S), aldolase (7.3S), and BSA (4.4S) (Pharmacia Biotech).

Mass spectrometry

Samples of MBP/gp21(338–425) in S-buffer (100 mg) were adjusted to pH 8.5 and then alkylated with 50 mM 4-vinylpyridine (Sigma Chemical Co., St. Louis, Missouri) for 30 min at room temperature, or reduced with 10 mM dithiothreitol for 3 h at 45°C and then alkylated. The samples were desalted and concentrated by precipitation in methanol/chloroform as described previously (Kobe et al., 1997) and then infused directly into a PE Sciex API III electrospray mass spectrometer in 15% (v/v) acetic acid, 50% (v/v) acetonitrile. The ion spectrum was visualized using the supplied software (Tune 2.5-FPU) and mass distributions calculated using the Hypermass facility in Mac-Spec 3.3 (Sciex).

Crystallization

Crystallization conditions were screened at 4°C and room temperature using the hanging drop-vapor diffusion technique (McPherson, 1982), by the sparse matrix approach (Jancarik & Kim, 1991) with commercially available screens (Hampton Research, Laguna Hills, California). The protein concentrations used in the initial trials were 10, 9.4, and 11 mg/mL for MBP/gp21(335–445), MBPL-/gp21(338–445), and MBPL-/gp21(338–425), respectively. Hanging drops comprised 2 μL of precipitant solution and 2 μL of protein sample inverted over 0.5 mL of reservoir solution.

X-ray diffraction data collection

Crystals were initially characterized using the MAR-Research image plate detector with CuK_α radiation from a Rigaku RU-200 rotating anode generator. Diffraction data were auto-indexed and processed with the programs DENZO and SCALEPACK (Otwinowski, 1993). High-resolution data for MBPL-/gp21(338–425) were collected at the EMBL BW7B beamline, DORIS storage ring, DESY, Hamburg ($\lambda = 0.8833 \text{ \AA}$) using a MAR-Research image plate detector. Prior to data collection, crystals were transiently soaked in the reservoir solution containing 40% ethylene glycol as cryoprotectant, and flash frozen at 100 K in a liquid nitrogen stream (Oxford Cryosystems).

Acknowledgments

We thank David Tribe, University of Melbourne, for supplying pDT1 DNA and P.L. Earl, National Institute of Allergy and Infectious Diseases, NIH, Maryland, for providing pPE16. Monoclonal antibody C8 was obtained through the AIDS Reference and Reagent Program, NIH, from G. Lewis. This study was supported by NHMRC Project Grant 960237. R.J.C. is a CARG scholar, and B.K. is a Wellcome Senior Research Fellow in Medical Science in Australia.

References

- Abacioglu YH, Fouts TR, Laman JD, Claassen E, Pincus SH, Moore JP, Roby CA, Kamin-Lewis R, Lewis GK. 1994. Epitope mapping and topology of

- baculovirus-expressed HIV-1 gp160 determined with a panel of murine monoclonal antibodies. *AIDS Res Hum Retroviruses* 10:371–381.
- Air GM, Webster RG, Colman PM, Laver WG. 1987. Distribution of sequence differences in influenza N9 neuraminidase of tern and whale viruses and crystallization of the whale neuraminidase complexed with antibodies. *Virology* 160:346–354.
- Blaber M, Zhang X-J, Lindstrom JD, Pepiot SD, Baase WA, Matthews BW. 1994. Determination of α -helical propensity within the context of a folded protein. *J Mol Biol* 235:600–624.
- Blacklow SC, Lu M, Kim PS. 1995. A trimeric subdomain of the simian immunodeficiency virus envelope glycoprotein. *Biochemistry* 34:14955–14962.
- Bullough PA, Hughson FM, Skehel JJ, Wiley DC. 1994. Structure of influenza hemagglutinin at the pH of membrane fusion. *Nature* 371:37–43.
- Carter DC, Rüker F, Ho JX, Lim K, Keeling K, Gilliland G, Ji X. 1994. Fusion proteins as alternate crystallization paths to difficult structure problems. *Protein Pept Lett* 1:175–178.
- CCP4. 1994. The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr D50*:760–763.
- Center RJ, Kemp BE, Pombourios P. 1997. Human immunodeficiency virus type 1 and 2 envelope glycoproteins oligomerize through conserved sequences. *J Virol* 71:5706–5711.
- Chan DC, Fass D, Berger JM, Kim PS. 1997. Core structure of gp41 from the HIV envelope glycoprotein. *Cell* 89:263–273.
- Delamarre L, Rosenberg AR, Pique C, Pham D, Callebaut I, Dokhlar M-C. 1996. The HTLV-1 envelope glycoproteins: Structure and functions. *J Acq Immun Defic Syndrome Hum Retrovirol* 13 Suppl 1:S85–S91.
- Delamarre L, Rosenberg AR, Pique C, Pham D, Dokhlar M-C. 1997. A novel human T-leukemia virus type 1 cell-to-cell transmission assay permits definition of SU glycoprotein amino acids important for infectivity. *J Virol* 71:259–266.
- Donahue JP, Patel H, Anderson WF, Hawiger J. 1994. Three-dimensional structure of the platelet integrin recognition segment of the fibrinogen γ chain obtained by carrier protein-driven crystallization. *Proc Natl Acad Sci USA* 91:12178–12182.
- Donegan E, Lee H, Operskalski GM, Shaw GM, Kleinman SH, Busch MP, Stevens CE, Schiff ER, Nowicki MJ, Hollingsworth CG, Mosley JW, Transfusion Safety Study Group. 1994. Transfusion transmission of retroviruses: Human T-lymphotropic virus types I and II compared with human immunodeficiency virus type 1. *Transfusion* 34:478–483.
- Earl PL, Hgin AW, Moss B. 1990. Removal of cryptic poxvirus transcription termination signals from the human immunodeficiency virus type 1 envelope gene enhances expression and immunogenicity of a recombinant vaccinia virus. *J Virol* 64:2448–2451.
- Einfeld DA, Hunter E. 1994. Expression of the TM protein of Rous sarcoma virus in the absence of SU shows that this domain is capable of oligomerization and intracellular transport. *J Virol* 68:2513–2520.
- Fan N, Gavalchin J, Paul B, Wells K, Lane MJ, Poiesz B. 1992. Infection of peripheral blood mononuclear cells and cell lines by cell-free human T cell lymphoma/leukemia virus type I. *J Clin Microbiol* 30:905–910.
- Fass D, Harrison SC, Kim PS. 1996. Retrovirus envelope domain at 1.7 resolution. *Nat Struct Biol* 3:465–469.
- Fass D, Kim PS. 1995. Dissection of a retrovirus envelope protein reveals structural similarity to influenza hemagglutinin. *Curr Biol* 5:1377–1383.
- Gallagher WR, Ball JM, Garry RF, Griffin MC, Montelaro RC. 1989. A general model for the transmembrane proteins of HIV and other retroviruses. *AIDS Res Hum Retroviruses* 5:431–440.
- Heinz DW, Baase WA, Matthews BW. 1992. Folding and function of a T4 lysozyme containing 10 consecutive alanines illustrate the redundancy of information in an amino acid sequence. *Proc Natl Acad Sci USA* 89:3751–3755.
- Jacobo-Molina A, Clark AD Jr, Williams RL, Nanni RG, Clark P, Ferris AL, Hughes SH, Arnold E. 1991. Crystals of a ternary complex of human immunodeficiency virus type 1 reverse transcriptase with a monoclonal antibody Fab fragment and double-stranded DNA diffract X-rays to 3.5- resolution. *Proc Natl Acad Sci USA* 88:10895–10899.
- Jancarik J, Kim S-H. 1991. Sparse matrix sampling: A screening method for crystallization of proteins. *J Appl Crystallogr* 24:409–411.
- Jones TA, Bergdoll M, Kjeldgaard M. 1990. O: A macromolecule modeling environment. In: Bugg CE, Ealick SE, eds. *Crystallographic and modeling methods in molecular design*. New York: Springer Verlag. pp 189–195.
- Kobe B, Jennings IG, House CM, Feil SC, Michell BJ, Tiganis T, Parker MW, Cotton RGH, Kemp BE. 1997. Regulation and crystallization of phosphorylated and dephosphorylated forms of truncated dimeric phenylalanine hydroxylase. *Protein Sci* 6:1352–1357.
- Kuge M, Fujii Y, Shimizu T, Hirose F, Matsukage A, Hakoshima T. 1997. Use of a fusion peptide to obtain crystals suitable for X-ray analysis: Crystallization of a GST-fused protein containing the DNA-binding domain of DNA replication-related element-binding factor, DREF. *Protein Sci* 6:1783–1786.
- Lim K, Ho JX, Keeling K, Gilliland GL, Ji X, Rker F, Carter DC. 1994. Three-dimensional structure of *Schistosoma japonicum* glutathione S-transferase fused with a six-amino acid conserved neutralizing epitope of gp41 from HIV. *Protein Sci* 3:2233–2244.
- Lu M, Blacklow SC, Kim PS. 1995. A trimeric structural domain of the HIV-1 transmembrane glycoprotein. *Nat Struct Biol* 2:1075–1082.
- Matthews BW. 1968. Solvent content of protein crystals. *J Mol Biol* 33:491–497.
- McPherson A. 1982. Preparation and analysis of protein crystals. New York: John Wiley and Sons.
- Ostermeier C, Iwata S, Ludwig B, Michel H. 1995. Fv fragment-mediated crystallization of the membrane protein bacterial cytochrome *c* oxidase. *Nat Struct Biol* 2:842–846.
- Otwinowski Z. 1993. Data collection and processing. In: Sawyer L, Isaacs N, Bailey S, eds. *Proceedings of the CCP4 study weekend*. Daresbury, UK: SERC Daresbury Laboratory. pp 56–62.
- Paine E, Garcia J, Philpott TC, Shaw G, Ratner L. 1991. Limited sequence variation in human T-lymphotropic virus type 1 isolates from North American and African patients. *Virology* 182:111–123.
- Paine E, Gu R, Ratner L. 1994. Structure and expression of the human T-cell leukemia virus type 1 envelope protein. *Virology* 199:331–338.
- Pancino G, Camoin L, Sonigo P. 1995. Structural analysis of the principal immunodominant domain of the feline immunodeficiency virus transmembrane glycoprotein. *J Virol* 69:2110–2118.
- Pique C, Pham D, Tursz T, Dokhlar M-C. 1992. Human T-cell leukemia virus type 1 envelope protein maturation process: Requirements for syncytium formation. *J Virol* 66:906–913.
- Pombourios P, Wilson KA, Center RJ, El Ahmar W, Kemp BE. 1997. Human immunodeficiency virus type 1 envelope glycoprotein oligomerization requires the gp41 amphipathic α -helical/leucine zipper-like sequence. *J Virol* 71:2041–2049.
- Prongay AJ, Smith TJ, Rossmann MG, Ehrlich LS, Carter CA, McClure J. 1990. Preparation and crystallization of a human immunodeficiency virus p24-Fab complex. *Proc Natl Acad Sci USA* 87:9980–9984.
- Rost B. 1996. PHD: Predicting one-dimensional protein structure by profiles-based neural networks. *Methods Enzymol* 266:525–559.
- Spurlino JC, Lu G-Y, Quicho FA. 1991. The 2.3- resolution structure of the maltose- or maltodextrin-binding protein, a primary receptor of bacterial active transport and chemotaxis. *J Biol Chem* 266:5202–5219.
- Syu W-J, Lee W-R, Du B, Yu Q-C, Essex M, Lee T-H. 1991. Role of conserved gp41 cysteine residues in the processing of human immunodeficiency virus envelope precursor and viral infectivity. *J Virol* 65:6349–6352.
- Tan K, Liu J-H, Wang J-H, Shen S, Lu M. 1997. Atomic structure of a thermostable subdomain of HIV-1 gp41. *Proc Natl Acad Sci USA* 94:12303–12308.
- Thomas A, Roth MJ. 1995. Analysis of cysteine mutations on the transmembrane protein of Moloney murine leukemia virus. *Virology* 211:285–289.
- Weissenhorn W, Calder LJ, Dessen A, Laue T, Skehel JJ, Wiley DC. 1997a. Assembly of a rod-shaped chimera of a trimeric GCN4 zipper and the HIV-1 gp41 ectodomain expressed in *Escherichia coli*. *Proc Natl Acad Sci USA* 94:6065–6069.
- Weissenhorn W, Dessen A, Harrison SC, Skehel JJ, Wiley DC. 1997b. Atomic structure of the ectodomain from HIV-1 gp41. *Nature* 387:426–430.
- Wilson IA, Skehel JJ, Wiley DC. 1981. Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 resolution. *Nature* 289:366–373.
- Wingfield PT, Stahl SJ, Kaufman J, Zlotnick A, Hyde CC, Gronenborn AM, Clore GM. 1997. The extracellular domain of immunodeficiency virus gp41 protein: Expression in *Escherichia coli*, purification, and crystallization. *Protein Sci* 6:1653–1660.