Biochemistry

 Volume 34, Number 46 November 21, 1995

Accelerated Publications

A Trimeric Subdomain of the Simian Immunodeficiency Virus Envelope Glycoprotein

Stephen C. Blacklow,[‡,§] Min Lu,[§] and Peter S. Kim*[,§]

Howard Hughes Medical Institute, Whitehead Institute for Biomedical Research, Department of Biology, Massachussetts Institute of Technology, Nine Cambridge Center, Cambridge, Massachusetts 02142, and Department of Pathology, Brigham & Women's Hospital, 75 Francis Street, Boston, Massachusetts 02115

Received August 4, 1995; Revised Manuscript Received October 5, 1995[⊗]

ABSTRACT: Previous attempts to define the oligomeric state of the HIV and SIV envelope glycoproteins have yielded conflicting results. We have produced in *Escherichia coli* a recombinant model for the ectodomain of the SIV envelope protein gp41 and have identified a small, trimeric subdomain by proteolytic digestion of this gp41 fragment. The subdomain assembles from two peptide fragments, spanning residues 28−80 (N_{28-80}) and residues 107−149 ($C_{107-149}$) of SIV gp41. Each of these peptides contains a 4,3-hydrophobic repeat, the hallmark of coiled-coil sequences. Upon mixing, the peptides form a highly helical, trimeric complex [3(N + C)] that resists proteolysis and has a melting temperature (T_m) above 90 °C in physiological buffer. The N- and C-terminal fragments are antiparallel to each other in the complex, as judged by the observation that digestion of a variant recombinant protein truncated at the amino terminus yields a C-terminal fragment shortened at its carboxy terminus. The N_{28-80} peptide contains more positions within the heptad repeat than $C_{107-149}$ that are predominantly hydrophobic, suggesting that N_{28-80} is buried in the interior of the complex. We propose that the complex consists of a parallel, trimeric coiled-coil of the N-terminal peptide, encircled by three C-terminal peptide helices arranged in an antiparallel fashion, and that this complex forms a core within the gp41 extracellular domain.

The immunodeficiency viruses HIV[1] and SIV attach to and gain entry into cells by means of an envelope glycoprotein (Env) that is present on the viral surface. Env is processed from a precursor, gp160, into a receptor-binding surface subunit, gp120, and a transmembrane subunit, gp41, necessary for fusion of the virus and host cell membranes [reviewed in Vainshav and Wong-Staal (1991)]. This functional organization of the viral envelope protein into two subunits, one for binding to the cellular receptor and the other essential for membrane fusion, is common to many viruses (Hunter & Swanstrom, 1990).

A high level of sequence variability is present in the Env proteins of the immunodeficiency viruses (Coffin, 1986). This sequence divergence is reflected in three significant phenotypic differences between laboratory strains of SIV and HIV. First, the HIV gp120 subunit binds CD4 more tightly than does the gp120 subunit of SIV Env (Sattentau et al., 1993). Second, the HIV gp120 subunit sheds from the virus upon binding of CD4 in some laboratory-passaged strains of HIV-1, whereas SIV gp120 does not, suggesting that the gp120 and gp41 subunits associate more tightly in SIV than in HIV (Moore et al., 1990; Sattenau & Moore, 1991; Sattentau et al., 1993). Finally, soluble CD4 inactivates laboratory-passaged strains of HIV (Moore et al., 1990), whereas it appears to enhance the ability of isolates of SIV

‡ Brigham and Women's Hospital.
§ Massachusetts Institute of Technology.
⊗ Abstract published in *Advance ACS Abstracts,* November 1, 1995.
[1] Abbreviations: HIV, human immunodeficiency virus; SIV, simian immunodeficiency virus; HPLC, high-performance liquid chromatography; IPTG, isopropyl β-D-thiogalactopyranoside; EDTA, (ethylenediamine)tetraacetic acid; DTT, dithiothreitol; DEAE, (diethylamino)-ethyl; PMSF, phenylmethanesulfonyl fluoride; CD, circular dichroism.

0006-2960/95/0434-14955$09.00/0

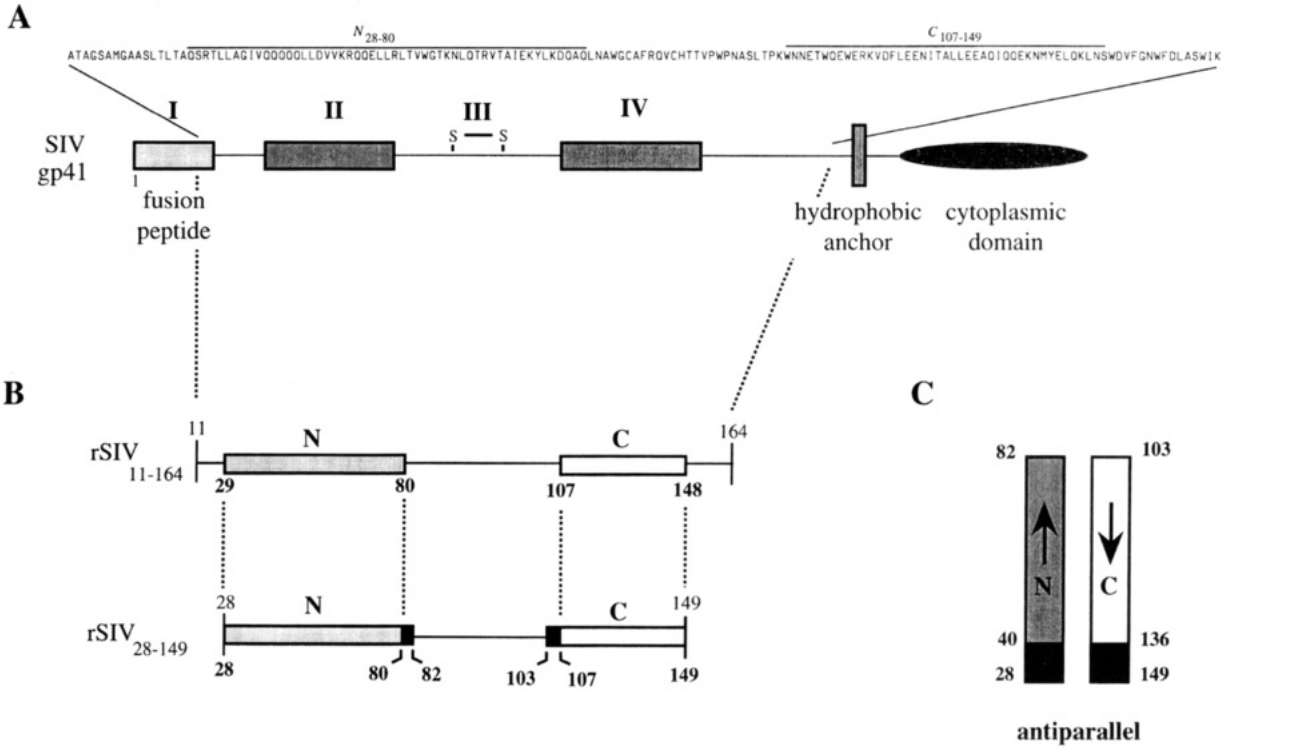

FIGURE 1: A. Schematic of gp41. The ectodomain consists of four distinct regions (I−IV). Regions II and IV contain 4,3-hydrophobic repeats. Sequence numbering is such that residue one is the first residue of SIV gp41 (residue 526 of Env). (B) Products of proteinase K digestion of the SIV_{11-164} and SIV_{28-149} recombinant proteins. Products from the N-terminal end of the protein are shaded gray, and products from the C-terminal end are in white. The termini of identified protease-resistant peptides are indicated in bold. (C) Comparison of products of proteinase K digestion of the SIV_{28-149} and SIV_{38-149} proteins, indicating that the N and C peptides are antiparallel in the N + C complex. The termini of identified protease-resistant peptides are indicated in bold.

(and of some primary isolates of HIV-1; Sullivan et al., 1995) to cause membrane fusion, a function in which the gp41 subunit plays the primary role (Allan et al., 1990).

The most conserved regions of Env lie within the ectodomain of gp41, and the ectodomains of gp41 from HIV and SIV isolates contain approximately 50−60% sequence identity. The striking similarities among the ectodomains of the envelope proteins from many different viruses [reviewed in Hunter and Swanstrom (1990)] have led to a working model for the structure of the extracellular domain of gp41 (Gallaher et al., 1989). This model contains four distinct regions (Figure 1). The first of these, the fusion peptide (Gallaher, 1987; Kowalski et al., 1987), is a hydrophobic, glycine-rich sequence that is necessary for the fusion of the viral and host cell membranes (Freed et al., 1990; Rafalski et al., 1990). The next segment contains a 4,3-hydrophobic repeat sequence (segment II of Figure 1A; Gallaher et al., 1989; Chambers et al., 1990; Delwart et al., 1990). Such 4,3-hydrophobic repeats ("leucine zippers") are characteristic of coiled-coil structures (Crick, 1953; Landschulz et al., 1988; O'Shea et al., 1989, 1991; Cohen & Parry, 1990), in which hydrophobic residues lie at the first (a) and fourth (d) positions within a seven amino acid repeat sequence. The third segment is a disulfide-bonded "loop" region, which is followed by the final region, a second 4,3-hydrophobic repeat (segment IV of Figure 1A) that precedes the transmembrane segment.

Protein dissection studies of the ectodomains of other viral envelope glycoproteins have led to insights about conformational changes that result in membrane fusion. The "spring-loaded" model for the conformational change of influenza hemagglutinin (HA) evolved from the observation that a "loop" in the crystal structure of the neutral-pH form of HA (Wilson et al., 1981) forms a coiled coil as an isolated peptide (Carr & Kim, 1993). The crystal structure of the low-pH form of the influenza HA2 subunit (disulfide bonded to a short peptide derived from the HA1 subunit) that confirms this model was solved from a protein fragment of the ectodomain of HA2 (Bullough et al., 1994). A model for the fusion-active form of the transmembrane subunit (p15) of Moloney murine leukemia virus has emerged from study of the ectodomain after removal of the hydrophobic, N-terminal fusion peptide (Fass & Kim, 1995).

By protein dissection, we recently identified a protease-resistant, trimeric core of the HIV gp41 ectodomain (Lu et al., 1995). The core consists of a complex of two discontinuous polypeptides from the envelope sequence, designated N and C. A model for the structure of the complex was proposed, on the basis of the biophysical properties and sequence characteristics of the two peptides (Lu et al., 1995).

Here, we have applied a similar protein dissection strategy to the ectodomain of SIV gp41. Despite 40−50% sequence divergence between the SIV and HIV gp41 ectodomains, a remarkably similar protease-resistant trimeric core, fully consistent with the proposed model, has been identified. The enhanced degree of sequence identity at the interior positions of the N and C 4,3-hydrophobic repeats of SIV and HIV Env likely accounts for the conservation of this structural motif.

MATERIALS AND METHODS

Materials. Plasmid $pSIVMac_{MFG}$, containing the gene encoding the envelope glycoprotein from SIV strain Mac239,

was a generous gift from D. Sanders and R. C. Mulligan (Whitehead Institute). *Escherichia coli* strain BL21(DE3) pLys(S) was used for expression of recombinant proteins (Studier et al., 1990). Proteinase K was purchased from Boeringer-Mannheim Biochemicals.

Recombinant Proteins. Cassettes encoding recombinant SIV ectodomain fragments $rSIV_{11-164}$, $rSIV_{28-149}$, and $rSIV_{38-149}$, flanked by *Nde*I and *Bam*HI restriction sites, were produced by the polymerase chain reaction (PCR) and inserted into the vector pAED4 (Doering, 1992). The identities of all amplified DNA fragments were verified by DNA sequencing. In $rSIV_{38-149}$, both cysteine residues were mutated to alanines to increase yield and simplify protein purification. Recombinant proteins were expressed in *E. coli* behind a T7 promoter (Studier et al., 1990), using the vector pAED4 (Doering, 1992). Cells were induced during logarithmic phase (OD_{600} of about 0.6) with IPTG and harvested after a 2-h induction period. Cells were lysed by sonication in 50 mM Tris buffer containing 25% sucrose (w/v) and 1 mM EDTA, pH 8.0. Recombinant proteins were purified from inclusion bodies and oxidized using published methods (Peng & Kim, 1994). The identities of the recombinant proteins were confirmed by N-terminal sequencing.

Limited Proteolysis Experiments. Purified rSIV recombinant proteins were subjected to proteolysis with proteinase K. rSIV proteins were suspended in 50 mM phosphate buffer containing 150 mM NaCl, pH 7.0. Initially, rSIV proteins were subjected to proteolysis at 37 °C for 2 h, using a 1:10 ratio of rSIV protein/protease (w/w). For large scale preparation of N and C peptides, digestion was performed for 2 h at 37 °C.

Proteinase K was inactivated by addition of PMSF to a final concentration of 3 mM, with further incubation at room temperature for 15 min. The proteolytic product was isolated by gel filtration chromatography on a G-75 Sephadex column. The N-peptide and C-peptide components of the protease-stable complexes were purified by reversed-phase HPLC on a Vydac C-18 column. Identities of the isolated peptides were established by N-terminal sequence analysis combined with laser desorption mass spectrometry (FinneganMAT, Lasermat). For quantitative reconstitution of the N_{28-80} + $C_{107-149}$ complex, lyophilized peptides were dissolved in water and mixed in equimolar amounts before dilution into 50 mM sodium phosphate buffer and 150 mM NaCl, pH 7.0. These proteolysis and reconstitution conditions were also used to evaluate the sensitivities of N_{28-80}, $C_{107-149}$, and the N_{28-80} + $C_{107-149}$ complex to proteolysis.

Biophysical Studies. Circular dichroism spectroscopy (CD) was performed on an Aviv 62DS spectrometer equipped with a thermoelectric temperature controller. Spectra were recorded at 0 °C in a 10 mm path length cuvette (1.5 nm bandwidth), using a 5 μM sample of protein in 50 mM sodium phosphate buffer and 150 mM NaCl, pH 7.0. Protein concentration was determined by the method of Edelhoch (1967) in 20 mM sodium phosphate buffer, pH 6.5, containing 6 M guanidinium chloride. The thermal dependence of the CD signal was monitored at 222 nm, using a 2-min equilibration period, a 30-s signal averaging time, and a step size of 2 °C. Samples for thermal melts consisted of 5 μM protein in 50 mM sodium phosphate buffer containing 150 mM NaCl, pH 7.0. The thermal denaturation curve of $C_{107-149}$ is >95% reversible; N_{28-80} precipitates after thermal denaturation. The thermal dependence of the CD signal of the N_{28-80} + $C_{107-149}$ complex is ~90% reversible at temperatures below 95 °C. An irreversible cooperative transition takes place with a midpoint at ~97 °C, and the recovered sample contains visible aggregates.

Sedimentation equilibrium measurements were performed at 20 °C on a Beckman Optima XL-A analytical ultracentrifuge, using an An-Ti rotor and six-sectored equilibrium centrifugation centerpieces. To determine the masses of the N_{28-80} + $C_{107-149}$ and N_{28-80} + $C_{103-149}$ complexes, rotor speeds of 13 000 and 16 000 rpm were used. Samples of 50 μM complex were prepared in 50 mM phosphate buffer containing 150 mM NaCl, pH 7.0, and were dialyzed exhaustively against the same buffer. Three dilutions of each sample (5−50 μM) were prepared by dilution with dialysate, and dialysate was used to fill the reference channels. N_{28-80} peptides were evaluated at concentrations of 15−150 μM and rotor speeds of 16 000 and 20 000 rpm. The apparent molecular weights were calculated by fitting data sets from each sector to a single ideal species model using the software program Kaleidagraph and partial specific volumes of 0.737 for the N_{28-80} + $C_{107-149}$ and N_{28-80} + $C_{103-149}$ complexes and 0.748 for the N_{28-80} peptide (Schuster & Laue, 1994).

RESULTS

Proteolytic Digestion of Recombinant gp41 Yields Two Peptide Fragments, N and C. A recombinant peptide model ($rSIV_{11-164}$) for the extraviral domain of SIV gp41, which lacks only 11 residues from the fusion peptide at the N-terminus of the ectodomain, was produced in *E. coli* (Figure 1B). Proteolytic digestion of $rSIV_{11-164}$ with proteinase K yields two major peptide fragments, identified by N-terminal sequencing and mass spectrometry (Figure 1B). These fragments correspond to two discontinuous segments of the polypeptide, spanning residues 29−80 (N_{29-80}), and 107−148 ($C_{107-148}$), with lengths of 52 and 41 residues, respectively.

Expression of $rSIV_{11-164}$ was limited to ~3 mg of protein per liter of cells. Remarkably, the $rSIV_{28-149}$ protein was expressed at ~50 mg per liter. $rSIV_{28-149}$ was digested with proteinase K, and the time course of digestion was monitored by reverse-phase HPLC. After ~10 min, the suspension had clarified, and two major peaks resistant to further digestion for at least 2 h emerged (data not shown). The major products of proteolysis, N_{28-80} and $C_{107-149}$, were almost identical to those obtained above (the fragments N_{28-82} and $C_{103-149}$ were also obtained; Figure 1B). The N_{28-80} and $C_{107-149}$ peptides were purified in larger quantities by HPLC for further characterization.

The N and C Peptides Form a Trimeric Complex. Neither N_{28-80} nor $C_{107-149}$ is resistant to proteolysis in isolation. Only when the N_{28-80} and $C_{107-149}$ peptides are mixed is a protease-resistant species formed (Figure 2). The stoichiometry of N_{28-80}:$C_{107-149}$ in the protected mixture is 1:1 (Figure 2). Equilibrium centrifugation of the reconstituted, protease-resistant species over a 10-fold range of concentration yields a mass of 34 900 Da, consistent with a composition for the complex of three N_{28-80}, and three $C_{107-149}$ peptides each (calculated mass 35 013; Figure 3).

The Complex is Fully Helical and Extremely Stable. The complex of N_{28-80} with $C_{107-149}$ is fully helical (Figure 4A).

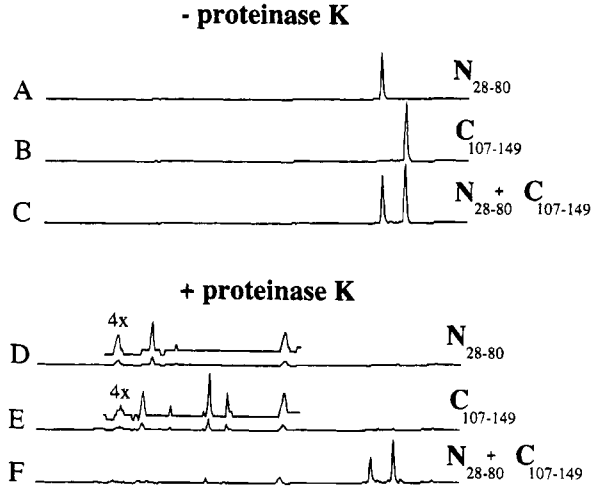

FIGURE 2: Reversed-phase HPLC chromatograms from the proteolysis protection experiments. (A) N_{28-80}, no protease. (B) $C_{107-149}$, no protease. (C) An equimolar mixture of N_{28-80} and $C_{107-149}$, no protease. (D) N_{28-80}, after proteolysis. (E) $C_{107-149}$, after proteolysis. (F) [N_{28-80} + $C_{107-149}$] after proteolysis.

The far-UV CD spectrum clearly differs from the weighted average of the spectra of the individual peptide components, N_{28-80} and $C_{107-149}$ (see below and Figure 4A), indicating formation of a complex. The complex has a melting temperature (T_m) above 90 °C (Figure 4B); this extraordinary stability to thermal denaturation likely explains the resistance of the complex to proteolysis.

The isolated N_{28-80} peptide is highly helical at neutral pH (Figure 4A). However, it does not form a unique oligomeric species, as judged by gel permeation chromatography (data not shown) and equilibrium centrifugation. The number of associated N_{28-80} peptides exceeds three at concentrations of 15 μM or greater (Figure 3C). The N_{28-80} oligomer undergoes a thermal unfolding transition with an apparent T_m of ~50 °C, but this transition is irreversible due to aggregation of the unfolded peptide (Figure 4B). The $C_{107-149}$ peptide, on the other hand, is unfolded in isolation (Figure 4A) and does not undergo a thermal transition (Figure 4B).

The N and C Peptides are Antiparallel in the Complex. The orientation of the N- and C-terminal peptides in the N + C complex was defined by proteinase K digestion of $rSIV_{38-149}$, which has a 10-residue N-terminal truncation (Figure 1C). Removal of the first 10 residues of N should expose the *first* 10 residues of C to proteolysis in a parallel complex or the *last* ten residues of C in an antiparallel complex. The major products of digestion of $rSIV_{38-149}$ are N_{40-82} (shortened by 13 residues at the N-terminus) and $C_{103-136}$ (trimmed by 13 residues at the C-terminus). These findings indicate that N and C associate antiparallel to each other in the complex (Figure 1C).

DISCUSSION

A Model for the Core Subdomain of gp41. Our results indicate that the SIV gp41 ectodomain contains a stable, protease-resistant core that spontaneously assembles from two components, N_{28-80} and $C_{107-149}$. This subdomain is a trimeric complex of these two components [3(N + C)]. The termini of this subdomain, defined by proteolysis of $rSIV_{11-164}$, are residue 29 (the amino terminus of N) and residue 148 (the carboxy terminus of C).

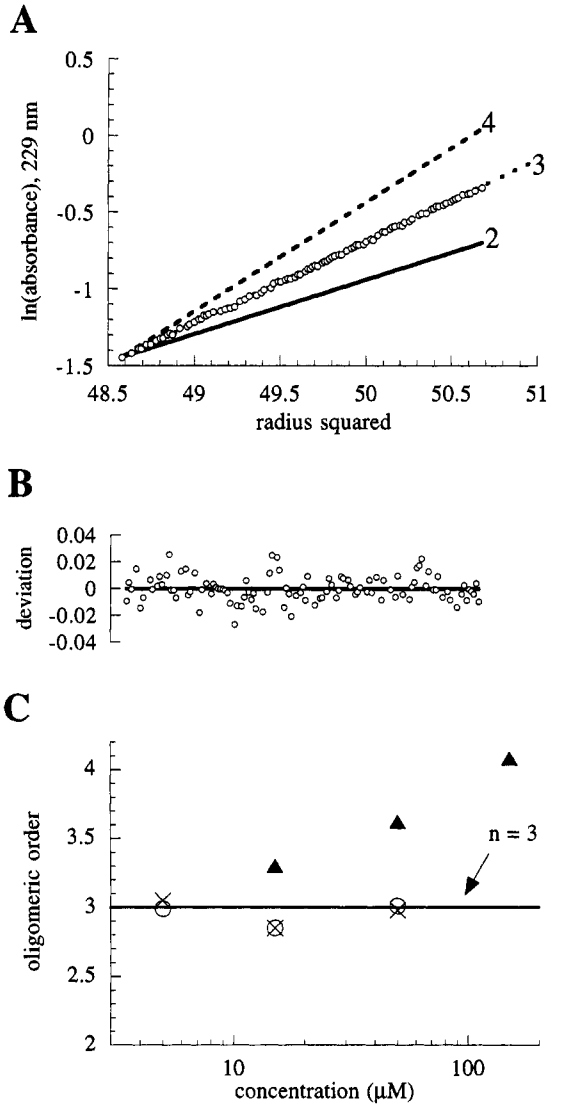

FIGURE 3: (A) Equilibrium centrifugation of the N_{28-80} + $C_{107-149}$ complex at 20 °C in physiological buffer (50 mM sodium phosphate, 150 mM NaCl, pH 7.0), 5 μM concentration. Log of the concentration distribution of protein as a function of the square of the radial position at 16 000 rpm and 20 °C. The best fit line describes a single homogeneous species of molecular mass 34 900 Da (ideal trimer model). Lines expected for dimeric and tetrameric complexes are included for comparison. (B) Residuals to the fit. No systematic trend is observed. (C) Oligomeric order (n) as a function of protein concentration for the N_{28-80} + $C_{107-149}$ complex (open circles), the N_{28-80} + $C_{103-149}$ complex (diagonal crosses), and the isolated N_{28-80} peptide (filled triangles). The horizontal line indicates the value n = 3.

We recently proposed a structural model for a similar N + C complex from the gp41 ectodomain of HIV, in which the N peptide forms an interior, trimeric coiled coil and associates with three surrounding C peptides aligned in an antiparallel orientation (Lu et al., 1995; Figure 6). The N_{28-80} + $C_{107-149}$ trimeric subdomain of the SIV gp41 ectodomain is fully consistent with this model, in which the hydrophobic e and g positions of adjacent N_{28-80} strands are buried in the protein interior, packed against the a and d hydrophobic residues of a single $C_{107-149}$ strand. The remaining residues, derived from the b, c, and f positions of N_{28-80} (see Figure 5A)[2] and the b, c, e, f, and g positions of $C_{107-149}$ (see Figure 5B), are more surface-exposed (Figure 5C), consistent with the predominance of polar and charged residues at these sites.

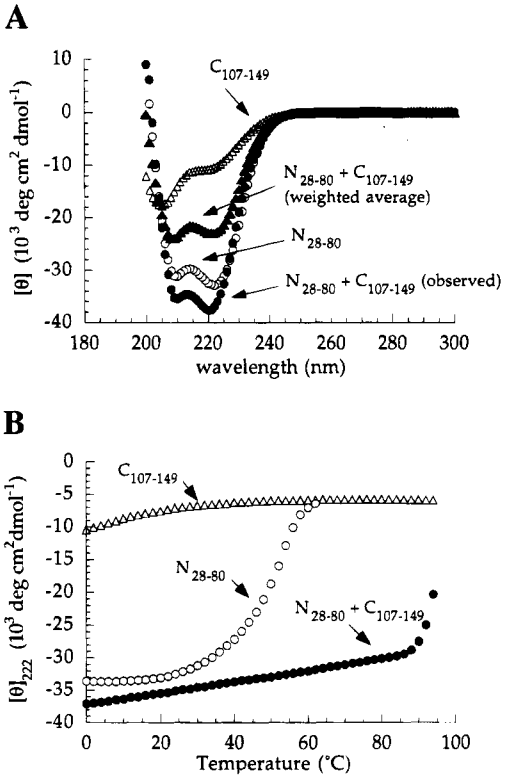

FIGURE 4: (A) Circular dichroism (CD) spectra of N_{28-80} (open circles), $C_{107-149}$ (open triangles), and the $N_{28-80} + C_{107-149}$ complex (filled circles). Spectra were recorded at 0 °C in physiological buffer (50 mM sodium phosphate, 150 mM NaCl, pH 7.0). The predicted spectrum for non-interacting $N_{28-80} + C_{107-149}$ peptides (filled triangles) is shown for comparison. (B) Thermal dependence of the CD signal at 222 nm in physiological buffer. Open circles, N_{28-80} peptide; open triangles, $C_{107-149}$ peptide; filled circles, $N_{28-80} + C_{107-149}$ complex.

The presence of both species, N_{28-80}, and $C_{107-149}$, is necessary for formation of a unique oligomer. Inspection of the helical wheel projection (Figure 5A) also suggests why the N_{28-80} peptide alone (and peptide DP-107, see below; Wild et al., 1992) does not form a unique oligomer. In coiled coils, the hydrophobic a and d positions of the heptad repeat form the core of the coiled-coil interface, and residues at the e and g positions pack against positions a and d to complete the hydrophobic interface. Two- and three-stranded coiled coils contain numerous interhelical salt bridges between charged side chains at the e and g positions, which are exposed substantially to solvent (O'Shea et al., 1991; Harbury et al., 1993; 1994). In contrast, the e and g positions of the GCN4-pLI *tetrameric* coiled coil are almost as buried as the a and d positions of the GCN4-p1 dimer (Harbury et al., 1993). In the N_{28-80} peptide, the a, d, e, and g positions are of similar hydrophobicity (Figure 5A), with few charged residues at the e and g positions. The extended hydrophobic surface in N_{28-80} may be prone to nonspecific aggregation when $C_{107-149}$ is not present.

Both HIV-1 and SIV gp41 have a similar cluster of four conserved glycosylation sites. Two of these sites lie between the N and C peptides, and two are within the sequence of the $C_{107-149}$ peptide (in SIV at N110, an e position, and N126, a g position). Site-directed mutagenesis of the asparagine residues of HIV-1 gp41 by several groups suggests that each of these sites is glycosylated *in vivo* (Lee et al., 1992; Fenouillet et al., 1994). No glycosylation sites are present in the N_{28-80} peptide. The location of these sites only within the $C_{107-149}$ peptide, which is proposed to be on the *exterior* of the N + C complex, is consistent with the general features of our model for the N + C complex.

The model of Figure 6 likely captures the essence of the structure of the N + C complex, although it is possible that glycosylation may lead to fraying of the amino-terminal end of the C-peptide and/or a register shift in the heptad repeat (Bullough et al., 1994). A register shift of the heptad repeat, for example, would place asparagine 126 at a more solvent-exposed position of the helix (a c position). It seems improbable, however, that proteolysis of the recombinant SIV and HIV-1 ectodomains of gp41 would artifactually generate such similar peptide fragments, and that the fragments from each viral source would spontaneously assemble into such stable, discrete, trimeric helical complexes.

Comparison with HIV Findings. Despite ~50% sequence divergence in the gp41 ectodomain between HIV and SIV, the precise boundaries of the proteinase K-resistant N and C peptides are essentially conserved (Figure 5C). The lengths of the two proteinase K resistant peptides are maintained, and the helical contents of the N + C complexes are comparable, strongly suggesting that the protease-resistant core represents a distinct stable structural subdomain.

The primary sequences of the HIV-1 and SIV N + C complexes show a striking pattern of sequence conservation, in which residues postulated to be buried in the interior of the N + C complex are strongly conserved (Figure 5C). In particular, the e and g positions of the N peptide, which comprise the interface for specific interaction with the C peptide in the structural model, are almost fully conserved (only a single A → V substitution at a g position exists). In addition, 10 of 15 residues at the a and d positions are identical, with a single nonconservative substitution (an I → T change) present at an a position. The a and d positions of the C peptide are also highly conserved (75% identity) between the SIV and HIV complexes. In contrast, positions postulated to be at the exterior of the model of the N + C complex have only 38% sequence identity and more frequent nonconservative substitutions.

The most notable difference between the SIV and HIV-1 subdomains is that observed in the CD spectra of the isolated N peptides. The N_{28-80} peptide from SIV has a characteristic, fully helical CD spectrum, whereas the N peptide from HIV gp41 has an atypical CD spectrum (Lu et al., 1995). Although it is not immediately apparent why the two spectra differ, one possibility is that T_{58} (an a position) may provide a buried polar interaction within the SIV N peptide that is

[2] We have evaluated the sequence of SIV gp41 with two algorithms used to predict coiled-coils. Remarkably, neither the program of Lupas et al. (1991) (maximum score, residues 28−55, of 1.14, <5% probability) nor the algorithm of Berger et al. (1995) (probability, residues 26−76 of <5%) predict that a sequence from region II forms a coiled coil. The program of Lupas et al. (1991) (maximum score, residues 113−140, of 1.74, >99% probability) predicts that a sequence from region IV forms a coiled coil, which is less strongly predicted by the algorithm of Berger et al. (1995) (probability, residues 110−138 of 26%). However, the database for the latter algorithm is limited to two-stranded coiled coils, and the Lupas et al. program is known to lead to a significant number of "false positives".

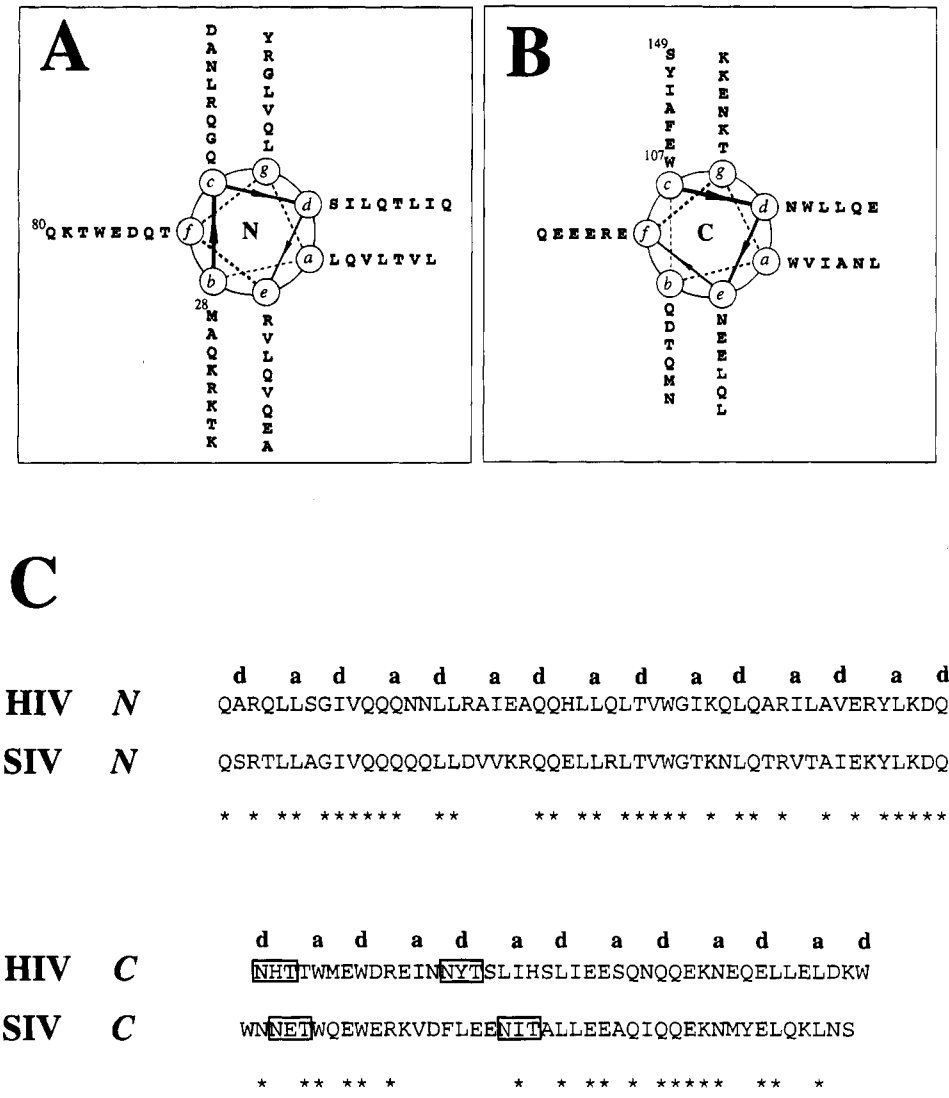

FIGURE 5: (A) Helical wheel projection of residues 28−80 of gp41 from SIV strain Mac239. A view down the helix axis from the amino terminus is shown. β-Branched amino acids prevail at the a and d positions of the heptad repeat, suggesting a trimeric oligomerization state (Harbury et al., 1993). The e and g positions are predominantly hydrophobic. (B) Helical wheel projection of residues 107−149 of gp41 from SIV strain Mac239. [see also footnote 2]. (C) Comparison of the primary sequences of the HIV-1_{Hxb2} and SIV_{Mac239} proteinase-K resistant N and C peptides. Positions of sequence identity are indicated by an asterisk (*). Heptad repeat positions are printed above in bold type. Conserved glycosylation sites are boxed.

not present in the HIV N peptide [the analogous residue in the HIV N peptide is isoleucine (Figure 5C)]. Buried polar interactions have been shown to be critical for maintaining orientational and oligomeric specificity in other coiled coils (Lumb & Kim, 1995).

Relationship to Previous Studies. Previous attempts to define the oligomeric state of the HIV and SIV envelope glycoproteins have yielded conflicting results. Findings consistent with dimeric, trimeric, and tetrameric forms of both HIV-1 Env (Gelderblom et al., 1989; Schawaller et al., 1989; Earl et al., 1990; Weiss et al., 1990; Doms et al., 1991) and SIV Env (Doms et al., 1990; Rey et al., 1990; Rhodes et al., 1994; Spies et al., 1994) have been reported. Our studies clearly indicate that the protease-resistant subdomains of both HIV-1 gp41 (Lu et al., 1995) and SIV gp41 are trimeric.

Our studies are also relevant for interpreting observations about two short peptides from the ectodomain of HIV gp41 that have antiviral activity (Wild et al., 1992, 1993, 1994; Chen et al., 1995). One of these, DP-107, which spans 38 residues of the N-terminal 4,3-hydrophobic repeat, is highly helical, but attempts to define its oligomeric state have yielded inconclusive results (Wild et al., 1992), perhaps for the reasons outlined above. A 36-residue peptide from the second (C-terminal) 4,3-hydrophobic sequence from HIV gp41, DP-178, is a more potent antiviral agent than DP-107 but lacks solution structure (Wild et al., 1994). When the DP-107 and DP-178 peptides are mixed, the resulting CD spectrum suggests association (Wild et al., 1995), and an antibody epitope has also been reconstituted by combining DP-178 with a fusion protein of gp41 linked to maltose binding protein (Chen et al., 1995). These results are consistent with our observation that the N_{28-80} and $C_{107-149}$ peptides of SIV gp41 combine in stoichiometric amounts to form a specific, trimeric, helical complex.

Relationship to Membrane Fusion Events. In the hemagglutinin molecule from influenza virus, the transmembrane subunit, HA2, undergoes a structural rearrangement after exposure to the mildly acidic endosome [for reviews, see Carr and Kim (1994) and Hughson (1995)]. The low-pH-

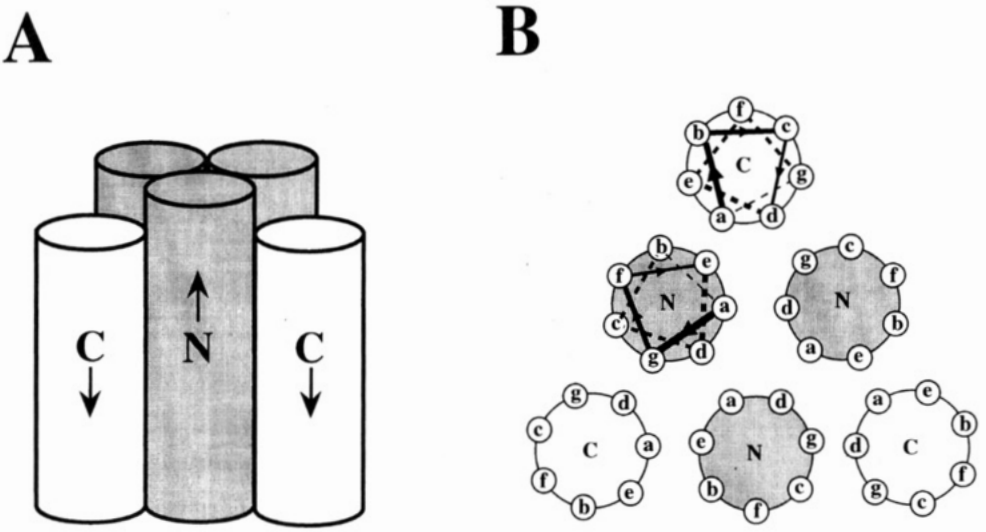

FIGURE 6: (A) Model of the N + C complex. This species is represented as a trimeric, parallel coiled coil of N_{28-80} in the center, encircled by three helices of $C_{107-149}$, oriented antiparallel to the N_{28-80} core. From the perspective shown, the rear C-helix is invisible. The precise vertical positioning of the shorter C-helix with respect to the N-helix is unknown. (B) Top view of the model presented in panel A, looking down the superhelical axis. Perspective (indicated by arrows and lines) is from the amino terminus of $C_{107-149}$ and from the carboxy terminus of N_{28-80}. Positions of the heptad repeat are indicated.

induced crystal structure (Bullough et al., 1994) of a proteolytic fragment of influenza hemagglutinin ($TBHA_2$) confirms the salient feature of a "spring-loaded" mechanism for the conformational change of influenza HA (Carr & Kim, 1993): a "loop" region of the native (neutral-pH) state (Wilson et al., 1981) is part of an extended coiled coil in the low-pH structure (Bullough et al., 1994), presumably repositioning the hydrophobic fusion peptide distally, in position to interact with the target membrane (Carr & Kim, 1993).

Numerous studies suggest that a conformational change also accompanies the conversion from the native to the fusogenic states in the HIV and SIV envelope glycoproteins (e.g., Moore et al., 1990, 1992; Clements et al., 1991; Sattenau et al., 1993), but the nature of this conformational change remains undefined. The subdomain identified in our studies may be present in the native state, the fusogenic state, or both.

The gp120 subunit sheds from the surface of virions after treatment with soluble CD4 in some strains of HIV-1 (Moore et al., 1990; Hart et al., 1991). Since the N + C complex assembles in the absence of gp120, one might argue that the complex models the fusogenic state of the transmembrane subunit gp41. This interpretation is also consistent with previous studies of the transmembrane subunits of influenza hemagglutinin (HA2) and Moloney murine leukemia virus (p15), which are both in the fusogenic conformation when their surface subunits, HA1 and gp70, are absent (Schoch et al., 1992; Carr & Kim, 1993; Bullough et al., 1994; Ragheb & Anderson, 1994; Fass & Kim, 1995). Moreover, the fusogenic conformations of both HA2 and p15 contain trimeric coiled coils adjacent to the fusion peptide (Carr & Kim, 1993; Bullough et al., 1994; Fan & Kim, 1995), and this organization is conserved in both HIV and SIV gp41, where the N_{28-80} trimeric coiled coil follows the fusion peptide [see also Chambers et al. (1990), Delwart et al. (1990), and Hunter and Swanstrom (1990)].

On the other hand, the N + C complex might represent a fragment of the native state of the protein, in which the fusion peptide is sequestered away from the target membrane. Only ~15 residues separate the last residue of $C_{107-149}$ from the viral membrane. If the complex is oriented perpendicular to the membrane, the antiparallel alignment of the N and C peptides would seem to direct the first residue of N_{28-80} toward the *viral* membrane not toward the *target* membrane. This topological problem might be avoided if the long axis of the protein is aligned parallel to the viral membrane, as seen in the crystal structure of tick-borne encephalitis virus (Rey et al., 1995). Although definitive evaluation of the gp41 ectodomain awaits high-resolution structural studies, the N + C subdomain of the SIV gp41 ectodomain described here is likely to play an important role in at least one state.

ACKNOWLEDGMENT

We thank Bonnie Berger for coiled-coil predictions, Chave Carr and Deborah Fass for helpful discussions, and Rheba Rutkowski and members of the Kim lab for critical reading of the manuscript. We are indebted to one reviewer for pointing out the locations of canonical glycosylation sites. Stephen C. Blacklow is a recipient of a Howard Hughes Medical Institute Postdoctoral Research Fellowship for Physicians.

REFERENCES

Allan, J. S., Strauss, J., & Buck, D. W. (1990) *Science 247*, 1084−1088.
Berger, B., Wilson, D. B., Tonchev, T., Milla, M., & Kim, P. S. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 8259−8263.
Bullough, P. A., Hughson, F. M., Skehel, J. J., & Wiley, D. C. (1994) *Nature 371*, 37−43.
Carr, C. M., & Kim, P. S. (1993) *Cell 73*, 823−832.
Carr, C. M., & Kim, P. S. (1994) *Science 266*, 234−236.
Chambers, P., Pringle, C. R., & Easton, A. J. (1990) *J. Gen. Virol. 71*, 3075−3080.
Chen, C. H., Matthews, T. J., McDanal, C. B., Bolognesi, D. P., & Greenberg, M. L. (1995) *J. Virol. 69*, 3771−3777.
Clements, G. J., Price-Jones, M. J., Stephens, P. E., Sutton, C., Schulz, T. F., Clapham, P. R., McKeating, J. A., McClure, M. O., Thomson, S., Marsh, M., Kay, J., Weiss, R. A., & Moore, J. P. (1991) *AIDS Res. Hum. Retroviruses 7*, 3−16.
Coffin, J. M. (1986) *Cell 46*, 1−4.
Cohen, C., & Parry, D. A. (1990) *Proteins 7*, 1−15.

Crick, F. H. C. (1953) *Acta Crystallogr. 6*, 685−689.
Delwart, E. L., Mosialos, G., & Gilmore, T. (1990) *AIDS Res. Hum. Retroviruses 6*, 703−706.
Doering, D. S. (1992) Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA.
Doms, R. W., Earl, P. L., Chakrabarti, S., & Moss, B. (1990) *J. Virol. 64*, 3537−3540.
Doms, R. W., Earl, P. L., & Moss, B. (1991) *Adv. Exp. Med. Biol. 300*, 203−219.
Earl, P. L., Doms, R. W., & Moss, B. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 648−652.
Edelhoch, H. (1967) *Biochemistry 6*, 1948−1954.
Fass, D., & Kim, P. S. (1995) *Current Biol.* (in press).
Fenouillet, E., Gluckman, J. C., & Jones, I. M. (1994) *Trends Biochem. Sci. 19*, 65−70.
Freed, E. O., Myers, D. J., & Risser, R. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 4650−4654.
Gallaher, W. R. (1987) *Cell 50*, 327−328.
Gallaher, W. R., Ball, J. M., Garry, R. F., Griffin, M. C., & Montelaro, R. C. (1989) *AIDS Res. Hum. Retroviruses 5*, 431−440.
Gelderblom, H. R., Ozel, M., & Pauli, G. (1989) *Arch. Virol. 106*, 1−13.
Harbury, P. B., Zhang, T., Kim, P. S., & Alber, T. (1993) *Science 262*, 1401−1407.
Harbury, P. B., Kim, P. S., & Alber, T. (1994) *Nature 371*, 80−83.
Hart, T. K., Kirsh, R., Ellens, H., Sweet, R. W., Lambert, D. M., Petteway, S. R., Jr., Leary, J., & Bugelski, P. J. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 2189−2193.
Hughson, F. M. (1995) *Curr. Biol. 5*, 265−274.
Hunter, E., & Swanstrom, R. (1990) *Curr. Top. Microbiol. Immunol. 157*, 187−253.
Kowalski, M., Potz, J., Basiripour, L., Dorfman, T., Goh, W. C., Terwilliger, E., Dayton, A., Rosen, C., Haseltine, W., & Sodroski, J. (1987) *Science 237*, 1351−1355.
Landschulz, W. H., Johnson, P. F., & McKnight, S. L. (1988) *Science 240*, 1759−1764.
Lee, W. R., Yu, X. F., Syu, W. J., Essex, M., & Lee, T. H. (1992) *J. Virol. 66*, 1799−1803.
Lu, M., Blacklow, S. C., & Kim, P. S. (1995) *Nature Struct. Biol.* (in press).
Lumb, K. J., & Kim, P. S. (1995) *Biochemistry 34*, 8642−8648.
Lupas, A., VanDyke, M., & Stock, J. (1991) *Science 252*, 1162−1164.
Moore, J. P., McKeating, J. A., Weiss, R. A., & Sattentau, Q. J. (1990) *Science 250*, 1139−1142.
Moore, J. P., Sattenau, Q. J., Klasse, P. J., & Burkly, L. C. (1992) *J. Virol. 66*, 4784−4793.
O'Shea, E. K., Rutkowski, R., & Kim, P. S. (1989) *Science 243*, 538−542.
O'Shea, E. K., Klemm, J. D., Kim, P. S., & Alber, T. (1991) *Science 254*, 539−544.
Peng, Z. Y., & Kim, P. S. (1994) *Biochemistry 33*, 2136−2141.
Rafalski, M., Lear, J. D., & De Grado, W. F. (1990) *Biochemistry 29*, 7917−7922.
Ragheb, J. A., & Anderson, W. F. (1994) *J. Virol. 68*, 3207−3219.
Rey, F. A., Heinz, F. X., Mandl, C., Kunz, C., & Harrison, S. C. (1995) *Nature 375*, 291−298.
Rey, M. A., Laurent, A. G., McClure, J., Krust, B., Montaigner, L., & Hovanessian, A. (1990) *J. Virol. 64*, 922−926.
Rhodes, A. D., Spitali, M., Hutchinson, G., Rud, E. W., & Stephens, P. E. (1994) *J. Gen. Virol. 75*, 207−213.
Sattenau, Q. J., & Moore, J. P. (1991) *J. Exp. Med. 174*, 407−415.
Sattenau, Q. J., Moore, J. P., Vignaux, F., Traincard, F., & Poignard, P. (1993) *J. Virol. 67*, 7383−7393.
Schawaller, M., Smith, G. E., Skehel, J. J., & Wiley, D. C. (1989) *Virology 172*, 367−369.
Schoch, C., Blumenthal, R., & Clague, M. J. (1992) *FEBS Lett. 311*, 221−225.
Schuster, T. M., & Laue, T. M. (1994) *Modern Analytical Ultracentrifugation*, p 351, Birkhäuser, Boston, MA.
Spies, C. P., Ritter, G., Jr., Mulligan, M. J., & Compans, R. W. (1994) *J. Virol. 68*, 585−591.
Studier, F. W., Rosenberg, A. H., Dunn, J. J., & Dubendorff, J. W. (1990) *Methods Enzymol. 185*, 60−89.
Sullivan, N., Sun, Y., Li, J., Hofmann, W., & Sodroski, J. (1995) *J. Virol. 69*, 4413−4422.
Vainshav, Y. N., & Wong-Staal, F. (1991) *Annu. Rev. Biochem. 60*, 577−630.
Weiss, C. D., Levy, J. A., & White, J. M. (1990) *J. Virol. 64*, 5674−5677.
Wild, C., Oas, T., McDanal, C., Bolognesi, D., & Matthews, T. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 10537−10541.
Wild, C., Greenwell, T., & Matthews, T. (1993) *AIDS Res. Hum. Retroviruses 9*, 1051−1053.
Wild, C. T., Shugars, D. C., Greenwell, T. K., McDanal, C. B., & Matthews, T. J. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 9770−9774.
Wild, C., Greenwell, T., Shugars, D., Rimsky-Clarke, L., & Matthews, T. (1995) *AIDS Res. Hum. Retroviruses 11*, 323−325.
Wilson, I. A., Skehel, J. J., & Wiley, D. C. (1981) *Nature 289*, 366−373.

BI951821X