The EMBO Journal Vol.17 No.16 pp.4572–4584, 1998

Three-dimensional solution structure of the 44 kDa ectodomain of SIV gp41

Michael Caffrey, Mengli Cai, Joshua Kaufman[1], Stephen J.Stahl[1], Paul T.Wingfield[1], David G.Covell[2], Angela M.Gronenborn[3] and G.Marius Clore[3]

Laboratory of Chemical Physics, Building 5, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892-0520, [1]Protein Expression Laboratory, Building 6B, National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health, Bethesda, MD 20892-2775 and [2]Frederick Cancer Research and Development Center, National Cancer Institute, Frederick, MD 21702, USA

[3]Corresponding authors
e-mail: clore@speck.niddk.nih.gov or gronenborn@vger.niddk.nih.gov

The solution structure of the ectodomain of simian immunodeficiency virus (SIV) gp41 (e-gp41), consisting of residues 27–149, has been determined by multidimensional heteronuclear NMR spectroscopy. SIV e-gp41 is a symmetric 44 kDa trimer with each subunit consisting of antiparallel N-terminal (residues 30–80) and C-terminal (residues 107–147) helices connected by a 26 residue loop (residues 81–106). The N-terminal helices of each subunit form a parallel coiled-coil structure in the interior of the complex which is surrounded by the C-terminal helices located on the exterior of the complex. The loop region is ordered and displays numerous intermolecular and non-sequential intramolecular contacts. The helical core of SIV e-gp41 is similar to recent X-ray structures of truncated constructs of the helical core of HIV-1 e-gp41. The present structure establishes unambiguously the connectivity of the N- and C-terminal helices in the trimer, and characterizes the conformation of the intervening loop, which has been implicated by mutagenesis and antibody epitope mapping to play a key role in gp120 association. In conjunction with previous studies, the solution structure of the SIV e-gp41 ectodomain provides insight into the binding site of gp120 and the mechanism of cell fusion. The present structure of SIV e-gp41 represents one of the largest protein structures determined by NMR to date.

Keywords: cell fusion/ectodomain/gp41/HIV/SIV

Introduction

The initial step of viral infection is often mediated by viral envelope proteins (Coffin, 1986; Freed and Martin, 1995). In the case of the human (HIV) and simian (SIV) immunodeficiency viruses, the envelope proteins exist as a complex of a surface subunit (gp120) and a transmembrane subunit (gp41), which are proteolytic products of the gp160 precursor coded by the *env* gene (Freed and Martin, 1995). During the initial step of HIV or SIV infection, the gp41–gp120 complex associates with the CD4 receptor and the chemokine coreceptor (Moore *et al*., 1993, 1997). Concurrently, gp41 dissociates from gp120, associates with the target membrane and mediates fusion of the viral and the cellular membranes by a process that involves the N-terminal hydrophobic region of gp41, termed the fusion peptide (Gallaher, 1987). Details of the fusion process, particularly concerning the association of gp41 and gp120, remain unknown, and structural information on the proteins involved is essential for the development of anti-HIV drugs.

gp41 comprises four functional domains: an N-terminal fusion peptide, an ectodomain (e-gp41), a transmembrane domain and a cytoplasmic domain (Freed and Martin, 1995). The ectodomain is the most conserved region of gp41, with 50–60% sequence identity between HIV-1 and SIV isolates (Douglas *et al*., 1997; Figure 1). The X-ray structure of subdomains of the HIV-1 e-gp41 have recently been solved (Chan *et al*., 1997; Tan *et al*., 1997; Weissenhorn *et al*., 1997). However, all three constructs employed in the crystallographic studies (Figure 1) omitted a central portion of e-gp41 comprising 35–44 residues, which had been shown by mutagenesis and antibody epitope mapping to play a role in gp120 association and possibly membrane fusion (Cao *et al*., 1993; Sattentau *et al*., 1993). In two recent papers we described the expression and purification (Wingfield *et al*., 1997), and determination of the secondary structure and global fold (Caffrey *et al*., 1997) of a soluble construct of trimeric SIV e-gp41 comprising residues 27–149. In this paper, we present the complete determination of the three-dimensional structure of the trimeric 44 kDa SIV e-gp41 by multidimensional nuclear magnetic resonance (NMR) spectroscopy.

Results and discussion

Structure determination

The solution structure of SIV e-gp41 was solved using double and triple resonance multidimensional NMR spectroscopy (Clore and Gronenborn, 1991, 1998; Bax and Grzesiek, 1993). The structure was determined on the basis of 2160 experimental NMR restraints per subunit, including 232 unambiguous intersubunit nuclear Overhauser enhancements (NOEs). Examples of the quality of the NMR data, showing a typical plane of a 4D $^{13}C/^{13}C$-separated NOE spectrum and strips from a 3D ^{13}C-separated/^{12}C-filtered NOE spectrum, are provided in Figure 2. Intermolecular NOEs were identified from various isotopically filtered NOE spectra recorded on 1:1 mixtures of $^{13}C/^{15}N/^{1}H$:$^{12}C/^{14}N/^{1}H$, $^{13}C/^{14}N/^{1}H$:$^{12}C/^{15}N/^{2}H$ and $^{13}C/^{15}N/^{2}H$:$^{12}C/^{14}N/^{1}H$ labeled SIV e-gp41, which enable one to specifically observe NOEs from protons attached to ^{13}C or ^{15}N to protons attached to ^{12}C or ^{14}N,

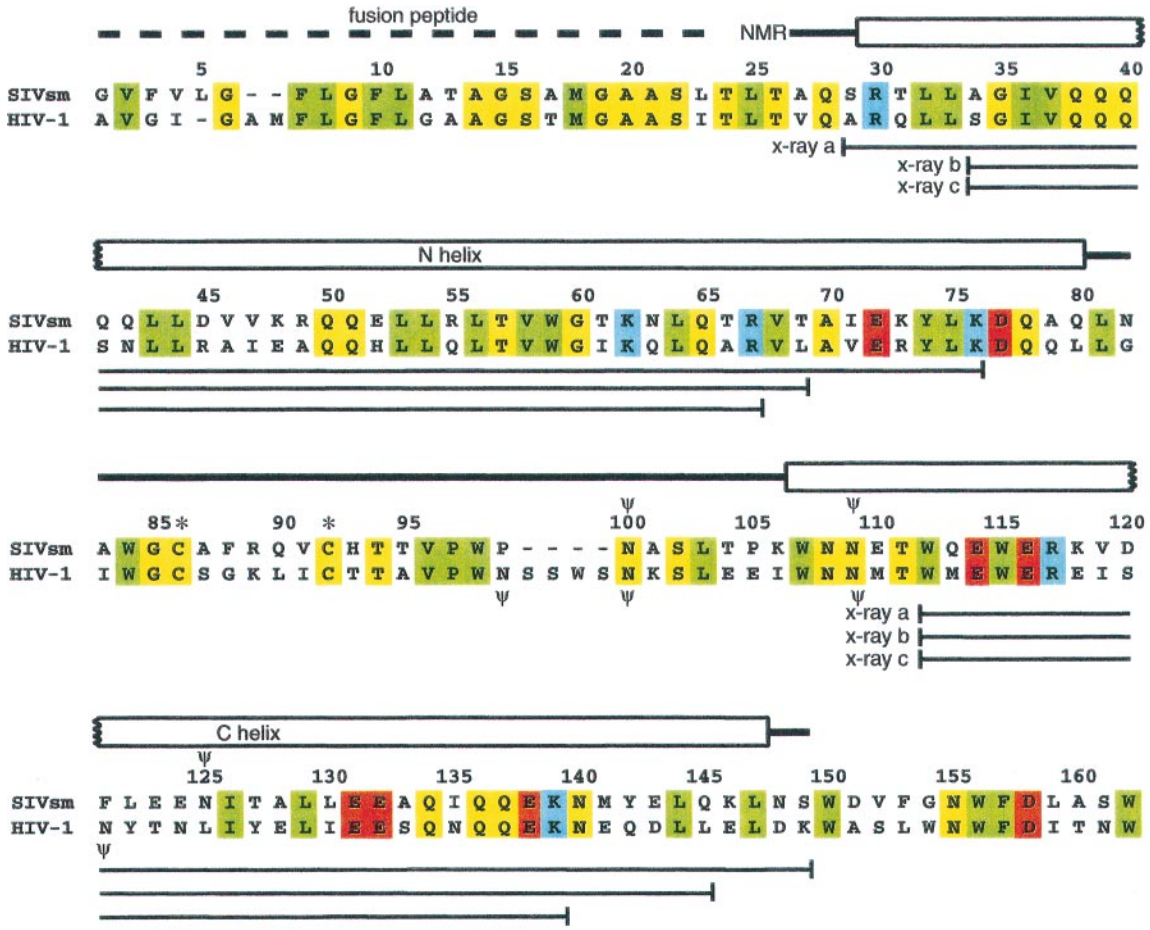

Fig. 1. Secondary structure of SIV e-gp41, and alignment of SIV (Sooty Mangabey) and HIV-1 (consensus sequence; Douglas *et al*., 1997) e-gp41. The numbering scheme employed is that of SIV e-gp41. Residues that are conserved in HIV-1 and SIV e-gp41 are color-coded as follows: green, hydrophobic; blue, positively charged; red, negatively charged; yellow, others. The location of the fusion peptide, the secondary structure and the construct used in the NMR study are indicated above the SIV sequence. The constructs used for the crystallographic studies are shown below the HIV-1 sequence and coded as follows: X-ray a, Weissenhorn *et al.* (1997); X-ray b, Chan *et al.* (1997); and X-ray c, Tan *et al*. (1997). The locations of putative glycosylation sites (Dedra *et al*., 1992) are indicated by the symbol ψ. The two cysteine residues within the loop, Cys86 and Cys92, are highlighted by an asterisk, and were mutated to alanine in the SIV e-gp41 construct used for NMR studies to prevent problems associated with multiple potential modes of intra- and intersubunit disulfide bond formation.

from protons attached to ^{13}C to protons attached to ^{15}N and from protons attached to ^{15}N to protons attached to ^{12}C or ^{14}N, respectively (Clore and Gronenborn, 1998). Since e-gp41 is trimeric, it is important to note that while it is possible to distinguish intra- from intersubunit NOEs using various types of heteronuclear filters, one cannot distinguish whether an intermolecular NOE arises from a close interproton distance contact between, for example, subunits A and B or subunits A and C. Thus, all intermolecular NOE restraints were treated as $(\Sigma r^{-6})^{-1/6}$ sums (Nilges, 1993), thereby permitting an NOE restraint from a specified proton on subunit A to be close to the specified proton on either subunit B or C, whichever is closest in the evolving calculated structures, the restraint being satisfied providing that at least one of the target protons is close. The superposition of the final 40 simulated annealing structures is shown in Figure 3 and the structural statistics are summarized in Table I.

Structure description

Ribbon diagrams and a molecular surface representation of SIV e-gp41 are displayed in Figure 4. The structure is a symmetric trimer and is cylindrical in shape, ~112 Å in length and ~35 Å in diameter, consistent with results from electron microscopy (Weissenhorn *et al*., 1996) and hydrodynamic modeling based on the observed sedimentation coefficient (Wingfield *et al*., 1997). The three subunits, which we term A (blue), B (red) and C (green), are arranged in a counter-clockwise manner when viewed from the top (Figure 4a, top panel) or side (Figure 4b) of the molecule. Each subunit comprises N-terminal (residues 30–80) and C-terminal (residues 107–147) helices, which are 81 and 62 Å long, respectively, connected by a long loop (residues 81–106) which protrudes ~26 Å upwards from the helical core. The contact surface (3726 $Å^2$ per subunit) between the subunits is extensive. The accessible surface area of the N-terminal helix, the loop and the C-terminal helix are reduced by 2434, 538 and 754 $Å^2$ per subunit, respectively, upon trimerization. Although the loop is somewhat more mobile than the helical core (Caffrey *et al*., 1997), it is still well-ordered as evidenced by numerous long-range and intermolecular NOEs (cf. Figure 2B). Indeed, within the loop we observed 128 medium-range ($1 < |i - j| \leq 4$) and 28 long-range

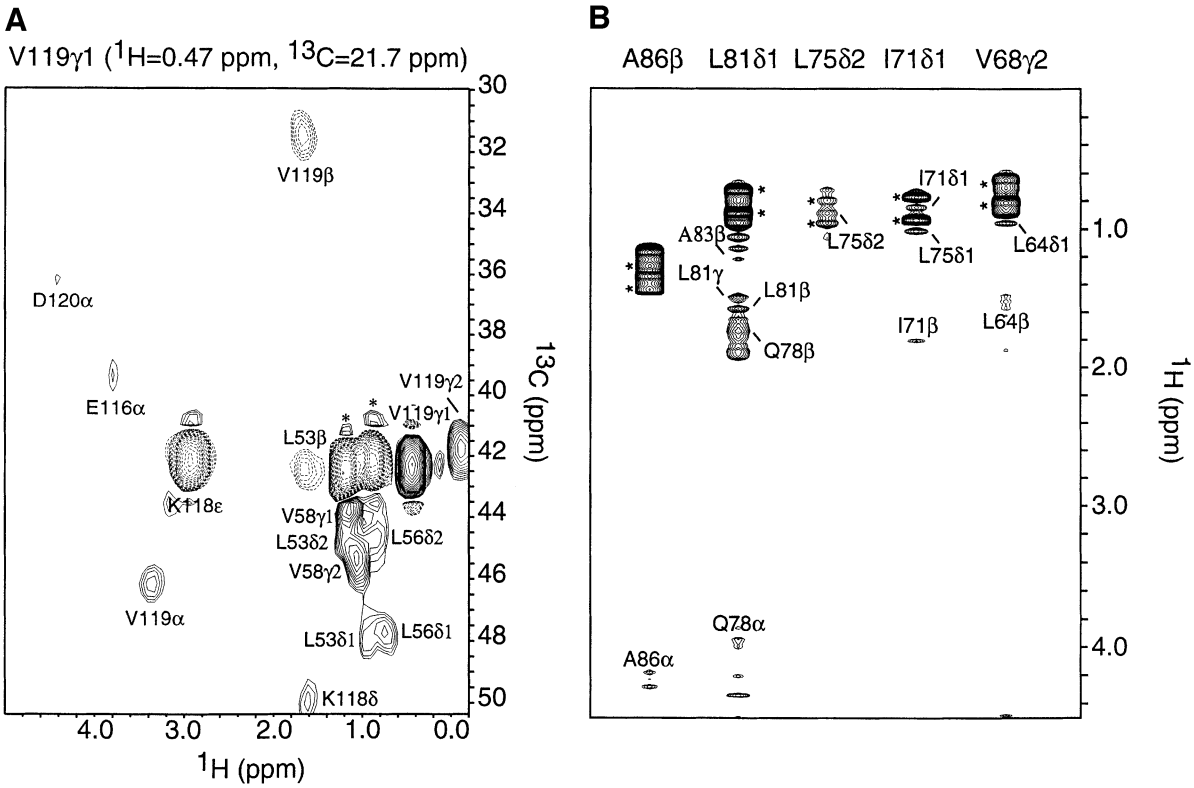

Fig. 2. (**A**) A ^{1}H-^{13}C plane taken from a 4D ^{13}C/^{13}C-edited NOE spectrum (80 ms mixing time) recorded on ^{13}C/^{15}N-labeled SIV e-gp41. The plane shown corresponds to that of Val119γ1. Cross-peaks which have their maximum intensity on other planes are denoted by asterisks. (**B**) Selected strips from a 3D ^{13}C-edited/^{12}C-filtered NOE spectrum (80 ms mixing time) recorded on a 1:1 mixture of ^{12}C/^{14}N- and ^{13}C/^{15}N-labeled SIV e-gp41, illustrating intersubunit contacts involving the loop (Leu81 and Ala86 strips) and the N-terminal helix (Val68, Ile71 and Leu75 strips). Residual diagonal peaks arising from ^{13}C-attached protons are denoted by asterisks.

(| i − j | >5) intramolecular NOEs per subunit and 24 intermolecular NOEs per subunit. Since intramolecular NOEs are readily distinguished from intermolecular ones, the connectivity between the N- and C-terminal helices (that is the assignment of the N- and C-terminal helices to individual subunits) is established unambigously in the NMR structure.

The N-terminal helices of the three subunits form a trimeric coiled-coil within the protein interior and are oriented parallel to each other at an angle of ~15°. The C-terminal helices are oriented antiparallel to the N-terminal helices, lying in the hydrophobic grooves formed by adjacent N-terminal helices and wrapping in a left-handed direction around the central coiled-coil. The C-terminal helices exhibit extensive intra- and intermolecular interactions with the N-terminal helices (contact surfaces ~925 and 754 Å^2, respectively). Intermolecular interactions exclusively involve contacts between the C-terminal helices of subunits A, B and C and the N-terminal helices of subunits C, A and B, respectively. The interhelical angles between the C-terminal helix of subunit A and the N-terminal helices of subunits A and C are ~165 and ~157°, respectively.

Figure 5 illustrates the distribution of amino acid types (hydrophobic, polar and other) in the structure, and the details of some of the internal sidechain packing between the subunits. There are extensive hydrophobic contacts between the N-terminal helices of the three subunits, and between the N and C-terminal helices. These include interactions involving 4–3 hydrophobic repeats of aliphatic residues such as Leu, Ile and Val (Figure 5c, f and g), as well as aromatic residues such as Trp (Figure 5d). In addition, there are a number of polar interactions between the helices. These include the intermolecular hydrogen-bonding network formed by the sidechain of Gln50 of the three subunits at the interface of the three N-terminal helices (Figure 5e), as well as a network of intermolecular hydrogen bonds between Gln38 of the N-terminal helix of subunits A, B and C, and Gln136 and Asn140 of the C-terminal helix of subunits B, C and A, respectively (Figure 5g). The loops of each subunit associate via hydrophobic interactions between their N-terminal portion, and include contacts involving Leu81, Ala83, Ala86, Ala87 and Phe88 (Figure 5b).

Surface hydrophobicity and electrostatic potential

Protein–protein interactions are largely dependent on hydrophobic contacts, supplemented by electrostatic interactions (Covell *et al.*, 1994; Young *et al.*, 1994; Jones and Thornton, 1996). We have therefore mapped the electrostatic potential (Figure 6a) and the two highest ranking surface hydrophobic clusters (computed as described by Young *et al.*, 1994) (Figure 6b) onto a molecular surface representation of e-gp41 in order to characterize potential binding sites for gp120.

The exposed surface of the C-terminal helix is predominantly negative, whereas that of the N-terminal helix is largely neutral, with the exception of two small patches of negative charge arising from Asp45 and Asp77, and two patches of positive charge, one formed by Arg49 and

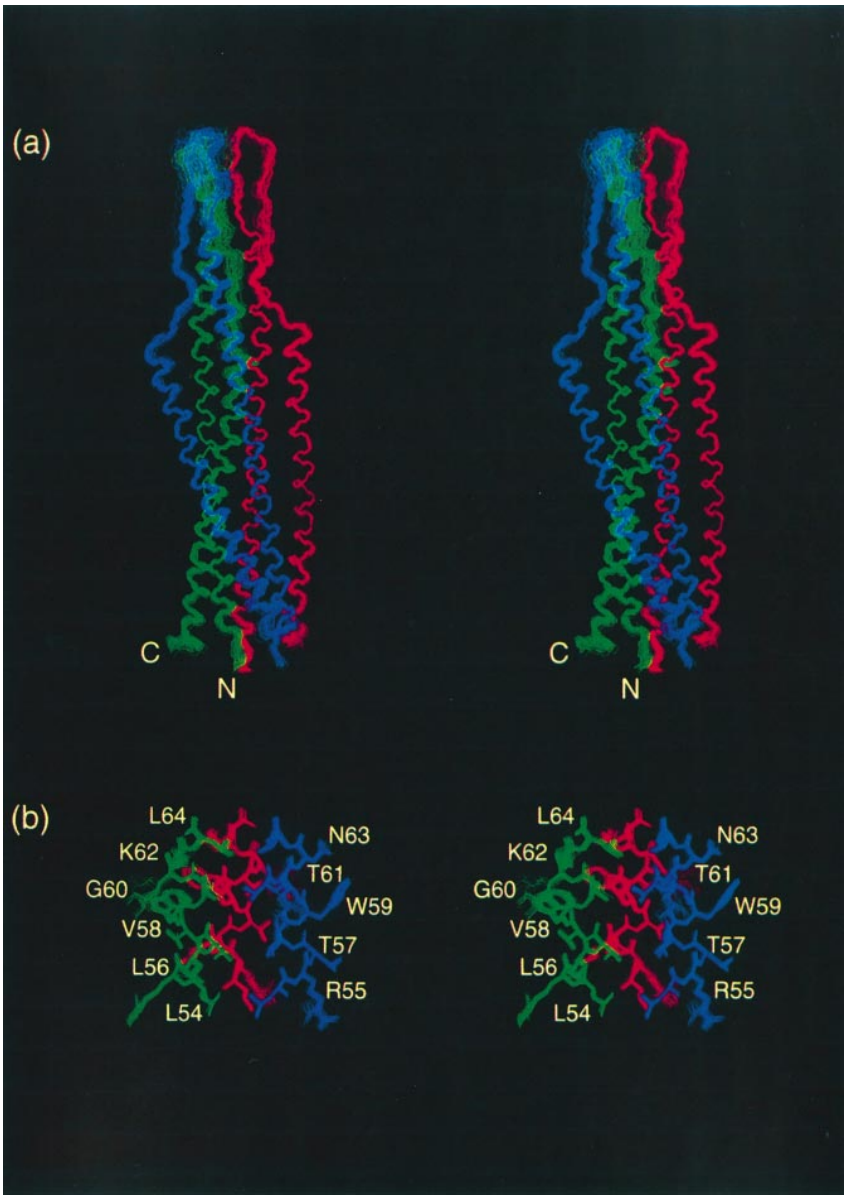

Fig. 3. Stereoviews showing best-fit superpositions of (**a**) the backbone (N, C_α, C′) and (**b**) selected side chains of the ensemble of 40 simulated annealing structures of SIV e-gp41. Subunits A, B and C are displayed in blue, red and green, respectively. The location of the N- and C-termini of subunit B are indicated in (a).

the other by Arg67. The loop, on the other hand, is mainly neutral, with the exception of its tip which bears a positive charge arising from Arg89.

The highest ranking hydrophobic cluster forms a cylinder that surrounds the lower two-thirds of the loop and the last turn of the N-terminal helix, comprises Asp77, Gln80, Trp84, Thr95, Val96, Pro97 and Trp98 of each subunit, and overlaps with the negative charge arising from Asp77. The second highest ranking hydrophobic cluster is located in the central region of the helical core of e-gp41 and comprises Asp45, Arg49, Glu52, Leu53, Leu56 and Trp59 of the N-terminal helix, and Lys118, Phe121 and Asn125 of the C-terminal helix. This hydrophobic cluster overlaps with two patches of negative charge, one arising from Glu116, Asp120, Glu123 and Glu124, the other from Asp45, and with a patch of positive charge arising from Arg49. Interestingly this region overlaps with the putative glycosylation site at Asn125 (Dedra *et al*., 1992).

We propose that the highest ranking hydrophobic cluster which involves the loop region may represent the binding surface for gp120, consistent with the results from mutagenesis (Cao *et al*., 1993) and antibody epitope mapping (Sattentau *et al*., 1993) studies (see below).

Correlation with biochemical studies

Within the loop region there is a conserved di-cysteine motif, which is common to the transmembrane subunit of all lentivirus surface envelope proteins and has been proposed to be important for gp120 association (Schulz *et al*., 1992). In the structure of SIV e-gp41 solved here, the two cysteines at positions 86 and 92 have been mutated to alanines to avoid multiple potential modes of intra- and intersubunit disulfide bond formation (Caffrey *et al*.,

Table I. Structural statistics[a]

	<SA>	($\overline{SA}$)r
r.m.s deviations from distance restraints[b]		
All (1500)	0.024 ± 0.001	0.023
Intramolecular		
Interresidue sequential (|i — j | = 1) (451)	0.015 ± 0.004	0.010
Interresidue short range (1 < | i — j| ⩽ 5) (368)	0.026 ± 0.002	0.030
Interresidue long range (|i — j | >5) (118)	0.042 ± 0.006	0.023
Intraresidue (153)[c]	0.008 ± 0.010	0.001
H-bonds (178)	0.023 ± 0.003	0.036
Intersubunit (232)[d]	0.028 ± 0.003	0.019
r.m.s. deviations from dihedral restraints (°) (360)[a]	0.286 ± 0.048	0.215
r.m.s. deviations from ^{3}J coupling constants[a]		
$^3J_{NHa}$ (Hz) (22)	1.09 ± 0.05	1.13
$^3J_{COCO}$ (Hz) (13)	0.73 ± 0.04	0.75
r.m.s. deviation from $^3DC_\alpha$(ND) isotope shift (ppb) (26)	1.04 ± 0.09	1.23
r.m.s. deviations from secondary ^{13}C shifts		
$^{13}C_\alpha$ (p.p.m.) (121)	1.23 ± 0.01	1.23
$^{13}C_\beta$ (p.p.m.) (118)	0.97 ± 0.02	0.96
Deviations from idealized covalent geometry		
bonds (Å) (2058)	0.003 ± 0.000	0.004
angles (°) (3722)	0.403 ± 0.009	0.585
impropers (°) (1079)	0.417 ± 0.035	0.645
Meaures of structure quality		
$E_{L\text{-}J}$ (kcal/mol)[e]	–1948 ± 21	–1682
PROCHECK[f]		
% residues in most favorable region of Ramachandran plot	90.5 ± 0.5	90.4
number of bad contacts/100 residues	6.9 ± 1.4	8.4
Coordinate precision of the trimer (Å)[g]		
backbone (N, C_α, C′, O) atoms	0.58 ± 0.16	
all heavy atoms	0.94 ± 0.18	

[a]The notation of the structures is as follows: <SA> are the final 40 simulated annealing structures; $\overline{SA}$ is the mean structure obtained by averaging the coordinates of the individual SA structures best-fitted to each other using residues 27–149 of the three subunits. ($\overline{SA}$)r is the restrained regularized mean structure obtained by restrained regularization of the mean structure $\overline{SA}$. The number of terms per subunit for the various restraints are given in parentheses. The final force constants employed for the various terms in the target function used for simulated annealing are as follows: 1000 kcal/mol/Å^2 for bond lengths; 500 kcal/mol/rad^2 for angles and improper torsions (which serve to maintain planarity and chirality); 4 kcal/mol/Å^4 for the quartic van der Waals repulsion term (with the van der Waals radii set to 0.8 times their value used in the CHARMM PARAM19/20 parameters); 50 kcal/mol/Å^2 for the non-crystallographic symmetry term; 30 kcal/mol/Å^2 for the experimental distance restraints (interproton distances and hydrogen bonds); 200 kcal/mol/rad^2 for the torsion angle restraints; 1 and 2 kcal/mol/Hz^2 for the $^3J_{HN\alpha}$ and $^3J_{COCO}$ coupling constant restraints, respectively; 0.5 kcal/mol/ppb^2 for the $^3\Delta C_\alpha$(ND) deuterium isotope shift restraints (which are related to the ψ backbone torsion angle); 0.5 kcal/mol/p.p.m.^{2} for the secondary ^{13}C chemical shift restraints; and 1.0 for the conformational database potential.
[b]None of the structures exhibited interproton distance violations >0.5 Å, dihedral angle violations >5°, or $^3J_{HN\alpha}$ or $^3J_{COCO}$ coupling constant violations >2 Hz. The torsion angle restraints consist of 119 ϕ, 120 ψ, 76 χ_1, 43 χ_2 and 2 χ_3 angles per subunit.
[c]Only structurally useful intraresidue NOEs are included in the restraints (i.e. involving protons separated by more than three bonds).
[d]These NOEs were treated as $(\Sigma r^{-6})^{-1/6}$ sums.
[e]The Lennard–Jones van der Waals energy was calculated with the CHARMM PARAM19/20 parameters and is not included in the target function for simulated annealing or restrained minimization.
[f]The overall quality of the structure was assessed using the program PROCHECK (Laskowski *et al*., 1993). There were no ϕ/ψ angles in the disallowed region of the Ramachandran plot. The dihedral angle G-factors for ϕ/ψ, χ_1/χ_2, χ_1, χ_3/χ_4 are 0.72 ± 0.02, 0.65 ± 0.05, –0.15 ± 0.12 and 0.11 ± 0.13, respectively.
[g]Defined as the average r.m.s. difference between the final 40 simulated annealing structures and the mean coordinates.

1997). These two Cys→Ala mutations have only a minimal effect on the structure since the ^{1}H–^{15}N correlation spectra of the wild-type and mutant SIV e-gp41 are virtually superimposable (Caffrey *et al*., 1997). The C_α–C_α separation between Ala86 and Ala92 within each subunit is ~9 Å, and between Ala86 of different subunits it is ~4 Å (Figures 4a and b, and 5b). This is consistent with the observation that both intra- and intersubunit disulfide bridges can be formed over time in wild-type HIV-1 and SIV e-gp41 (Weissenhorn *et al*., 1996; Wingfield *et al*., 1997).

There are three putative N-glycosylation sites in SIV e-gp41, one in the loop (Asn100) and two in the C-terminal helix (Asn109 and Asn125) (Dedra *et al*., 1992) (Figure 4c). All three asparagines are located on the exterior of the structure. Finally, the binding site for the neutralizing antibody 2F5 (Muster *et al*., 1993) occurs at the end of the C-terminal helix in a region which is solvent exposed and proximal to the fusion peptide (Figure 4c). This epitope is known to be exposed in the presence of gp120 (Satteneau *et al*., 1995), indicating that it cannot overlap with the gp120 binding site. This is fully consistent with the proposed location of the gp120 binding site in the loop region (see above).

Comparison with HIV-1 gp41

There is extensive sequence similarity (with 56% overall sequence identity) throughout e-gp41 of HIV-1 and SIV (Figure 1). The major differences correspond to a four residue deletion in the loop region of SIV e-gp41 relative to that of HIV-1 e-gp41, and a generally lower degree of sequence identity in the loop region and C-terminal helix. Specifically, the extents of sequence identity for the N-terminal helix, loop region and C-terminal helix are 64,

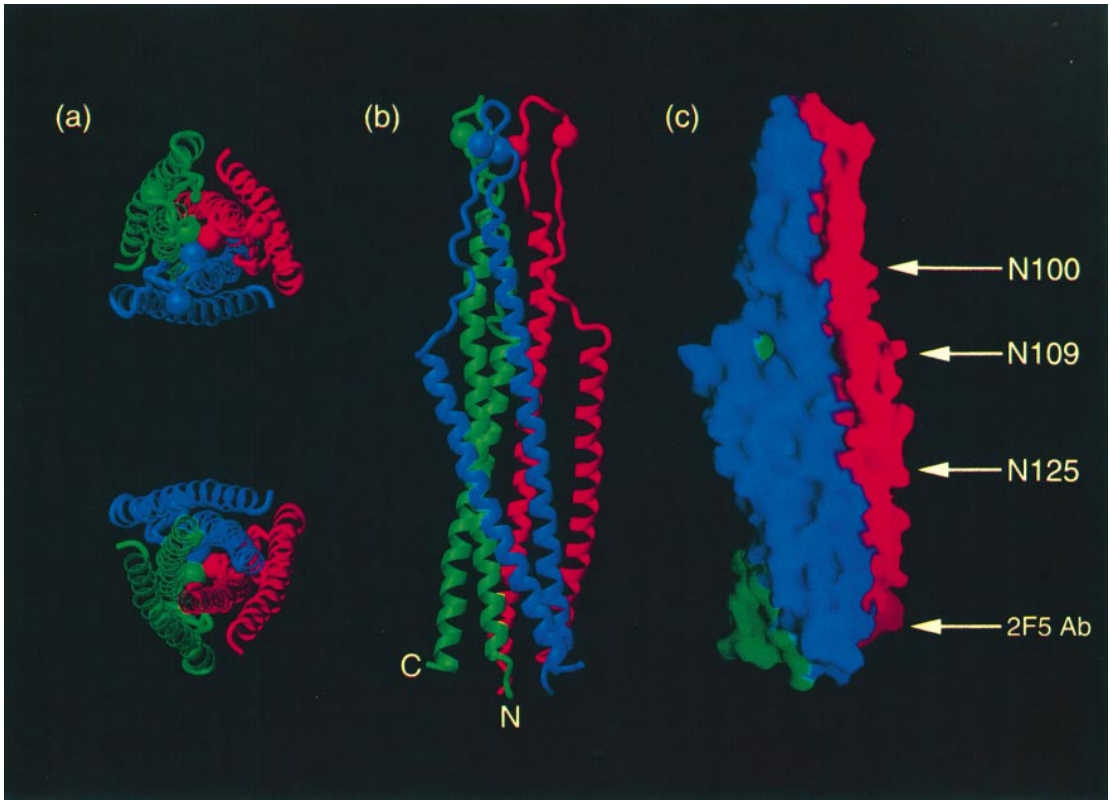

Fig. 4. Ribbon diagrams of SIV e-gp41 viewed from (**a**) the top (upper panel) and bottom (lower panel) and (**b**) the side with respect to the viral membrane (note that the transmembrane segment of gp41 is connected to the C-terminal helices via a flexible 20 residue linker). (**c**) Molecular surface of e-gp41 displayed in the same orientation as the ribbon diagram shown in (b). Subunits A, B and C are displayed in blue, red and green, respectively. The location in the loop of the C_α atoms of residues 86 and 92 which are cysteines in the wild-type sequence but have been mutated to alanine in the current structure are indicated by spheres in (a) and (b). The locations of the three putative glycosylation sites (Dedra *et al*., 1992) and the epitope for the 2F5 neutralizing monoclonal antibody (Muster *et al*., 1993) are indicated in (c).

46 and 53%, respectively. Figure 7 provides a comparison of the NMR structure of SIV e-gp41 (residues 27–149) with structures of truncated versions of the helical core of HIV-1 gp41 determined by X-ray crystallography. The three X-ray structures of HIV-1 e-gp41 correspond to residues 34–67 and 112–139 (Tan *et al*., 1997), 34–69 and 112–145 (Chan *et al*., 1997), and 29–76 and 112–149 (Weissenhorn *et al*., 1997) using the SIV numbering scheme, and are 54, 58 and 73 Å in length, respectively, compared with 112 Å for the complete e-gp41 of SIV. The central helical core of SIV e-gp41 is structurally very similar to that of the HIV-1 peptides in agreement with the high degree of sequence identity between HIV-1 and SIV proteins. The backbone of the X-ray structures of Tan *et al*. (1997), Chan *et al*. (1997) and Weissenhorn *et al*. (1997) can be superimposed onto the corresponding regions of SIV e-gp41 with backbone atomic r.m.s. differences of 0.5, 0.8 and 0.8 Å, respectively. Note that the N- and C-terminal helices of SIV e-gp41 clearly extend 4 and 5 residues, repectively (SIV residues 77–80 and 107–111), with respect to the most complete X-ray structure of HIV-1 e-gp41 (Weissenhorn *et al*., 1997). Most importantly, the present structure includes the central 35 residues of the ectodomain, which have been implicated previously in gp120 association and membrane fusion (discussed below). Finally, the present NMR structure of SIV e-gp41 establishes unambiguously the connectivity of the N- and C-terminal helices, which previously could only be inferred by the limited functional and structural homology to influenza virus hemagglutinin (Weissenhorn *et al*., 1997). This connectivity coincides with that observed in the X-ray structure of Tan *et al.* (1997), where the portions of the N-terminal (residues 34–67) and C-terminal (residues 112–139) helices employed in the construct were artificially connected by a six residue linker. It should be noted, however, that this short linker only permitted one connectivity, whereas several permutations are potentially feasible when the two helices are connected by the long 26 residue loop.

Correlation with mutagenesis studies

A number of site-directed mutants of HIV-1 e-gp41 have been studied (Cao *et al*., 1993; Chen, 1994). One set of mutations severely impairs the processing of gp160 into gp120 and gp41 and hence automatically results in an apparent decrease in both association with gp120 and cell fusion in the *in vivo* assays employed. The other set of mutations, for which processing exceeds ~20%, can be divided into three classes which are displayed in Figure 6c. All these mutations involve residues that are conserved in HIV-1 and SIV e-gp41.

The first class of mutations, green in Figure 6c, has no significant effect on either gp120 association or fusion, and is exclusively located on the exposed surface of the C-terminal helix (Glu131→Leu, Gln136→Leu, Leu147→Phe) close to the base of e-gp41, suggesting that these residues are not involved in either gp120 binding or membrane fusion. While Glu131 is completely solvent

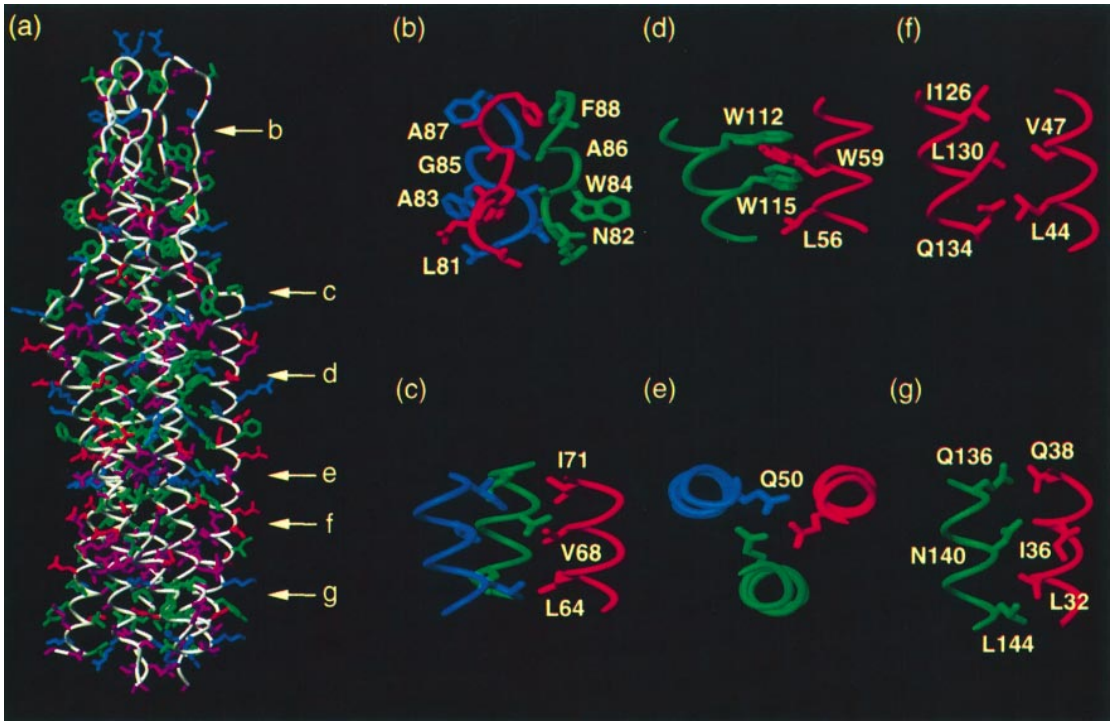

Fig. 5. (**a**) Overall view of SIV e-gp41 illustrating the distribution of amino acid types with the backbone shown as a C_α worm (white), and hydrophobic (aliphatic and aromatic), positively charged (Arg, Lys and His), negatively charged (Asp and Glu) and other amino acids displayed in green, blue, red and magenta, respectively. Sidechain interactions illustrating (**b**) intermolecular contacts between the loops, (**c** and **e**) intermolecular contacts between the N-terminal helices, (**d** and **g**) intermolecular contacts between the N- and C-terminal helices, and (**f**) intramolecular contacts between the N- and C-terminal helices; subunits A, B and C are color-coded blue, red and green, respectively. The location of (b) to (g) in the full structure is indicated in (a).

accessible, the other two residues, Gln136 and Leu147, are only partially solvent accessible and do participate in some intersubunit interactions between the C- and N-terminal helices. The Leu147→Phe mutation, however, is conservative in nature and would be expected to be well tolerated. Gln136(A), on the other hand, is involved in an intersubunit hydrogen bonding network with Gln38(C) and Gln42(C) (Figure 5g); presumably replacement of Gln136 by Leu, while removing these intermolecular hydrogen bonds preserves and possibly enhances the intersubunit hydrophobic contacts sufficiently to have little or no effect on the stability or structure of e-gp41. These results also suggest that gp120 does not bind to this region of gp41, consistent with the model presented above.

The second class of mutations, red in Figure 6c, significantly reduces or abolishes cell fusion but has little impact on gp120 association; all these mutations (with the exception of two, Asp77→Leu and Trp84→Met) comprise residues that participate in intersubunit contacts and would be expected to destabilize the trimer (that is to shift the monomer–trimer equilibrium towards the monomeric form). Mutants Ile36→Ala, Gln50→Leu (Figure 5e), Leu54→Gly or Pro and Leu64→Pro (Figure 5c) involve substitutions at the interface of the three subunits formed by the three N-terminal helices. Mutants Leu56→Ala or Pro (Figure 5d), Trp59→Arg (Figure 5d) and Asn140→Leu (Figure 5g) involve substitutions at points of intersubunit contact between the N- and C-terminal helices. The Leu53→Pro mutation involves a residue that participates in intermolecular contacts between both the N-terminal helices, and between the N- and C-terminal helices. Both Asp77 and Trp84 are located in the loop, are accessible to solvent and form part of the highest ranking surface hydrophobic cluster (Figure 6b). This suggests that one possible explanation for the effects of the Asp77→Leu and Trp84→Met mutations is that they may strengthen the association between gp41 and gp120 and thus prevent fusion by inhibiting dissociation of gp120.

Finally, the third class of mutations, yellow in Figure 6c, abolishes (Val96→Ser) or significantly reduces (Leu75→Pro) gp120 association, and, as a consequence, fusion can no longer be initiated. Val96 is located in the loop and is completely solvent accessible, suggesting that it may be a critical residue involved in gp120 association. Interestingly, Val96 is located in the highest ranking surface hydrophobic cluster on e-gp41 (Figure 6b). Leu75, on the other hand, is located at the C-terminus of the N-terminal helix and participates in intersubunit contacts at the trimer interface. The Leu75→Pro substitution may result in premature termination of the N-terminal helix and distort the subsequent conformation of the loop, thereby disrupting the gp120 binding site.

gp120 binding site

Taken as a whole, the mutational and structural data suggest that the stability of the trimer is more critical to cell fusion than to gp120 association and that the surface hydrophobic cluster on the loop presents the most likely binding site for gp120 (cf. Figure 6b). This is consistent with the finding that in HIV-1 dissociation of gp120 from gp41 exposes an epitope, recognized by the murine

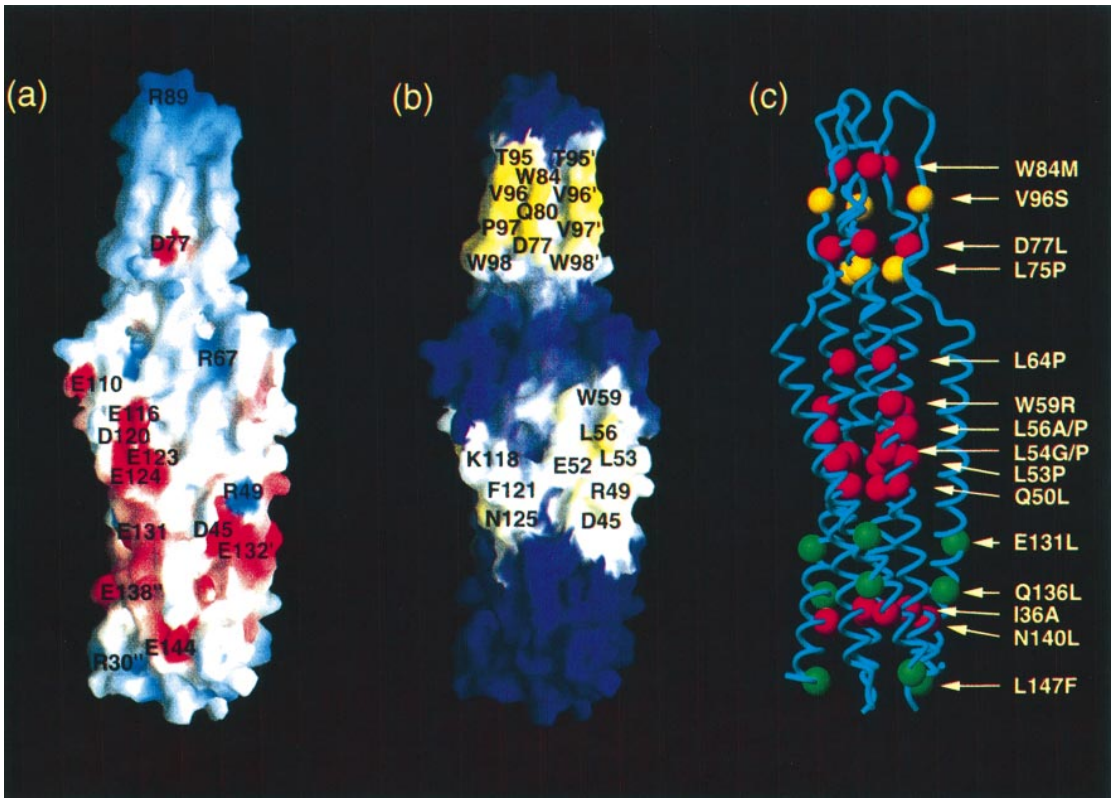

Fig. 6. Mapping of (**a**) the electrostatic potential and (**b**) the two highest ranking surface hydrophobic clusters on the molecular surface of SIV e-gp41. The electrostatic potential is colored from red (negative charge) to blue (positive charge); regions of highest hydrophobicity are yellow, those of lowest hydrophobicity are purple, and the gradient from yellow to white to purple corresponds to decreasing hydrophobicity; residues of subunits B and C are denoted by single and double apostrophes, respectively. (**c**) Ribbon diagram of SIV e-gp41 in the same orientation as the molecular surfaces shown in (a) and (b) illustrating the distribution of mutations. The mutations are color-coded as follows: yellow, reduces or abolishes gp120 binding and abolishes fusion; red, abolishes fusion but not gp120 binding; and green, has no significant effect on either gp120 binding or fusion.

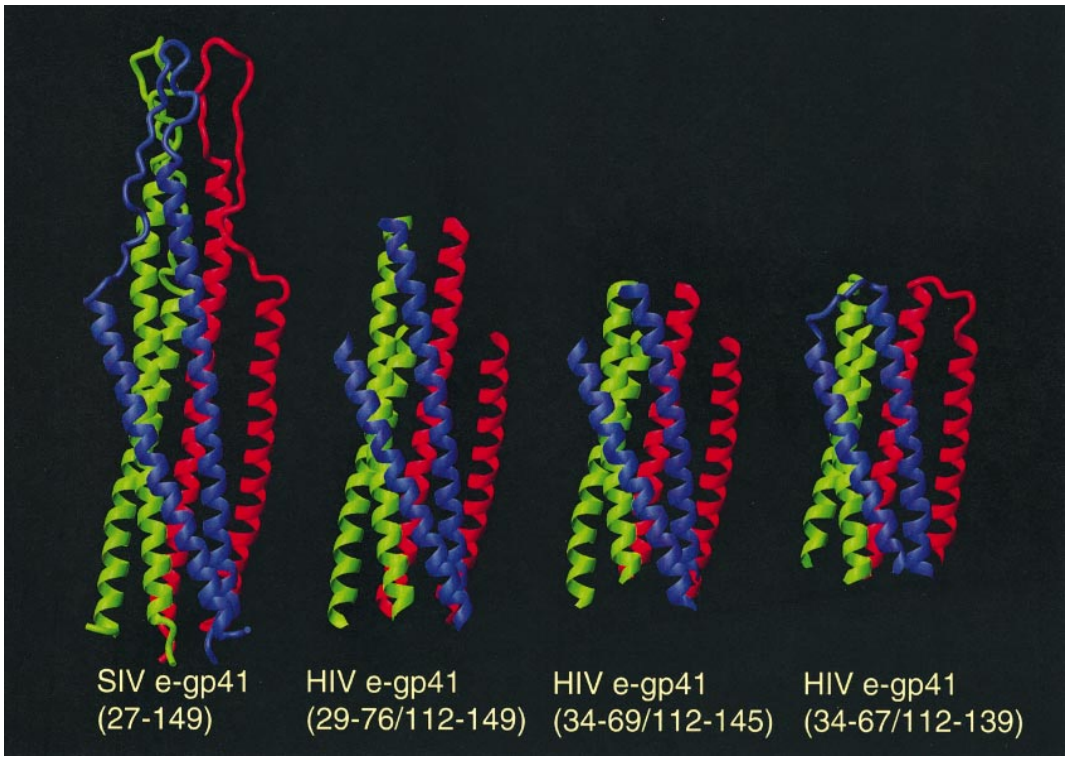

Fig. 7. Comparison of the solution structure of SIV e-gp41 with the X-ray structures of truncated versions of HIV-1 e-gp41. The residue numbering corresponds to that of SIV gp41 (see Figure 1). The X-ray structures were taken from Weissenhorn *et al*. (1997), Chan *et al*. (1997) and Tan *et al*. (1997) (from left to right). All structures are viewed in the same orientation.

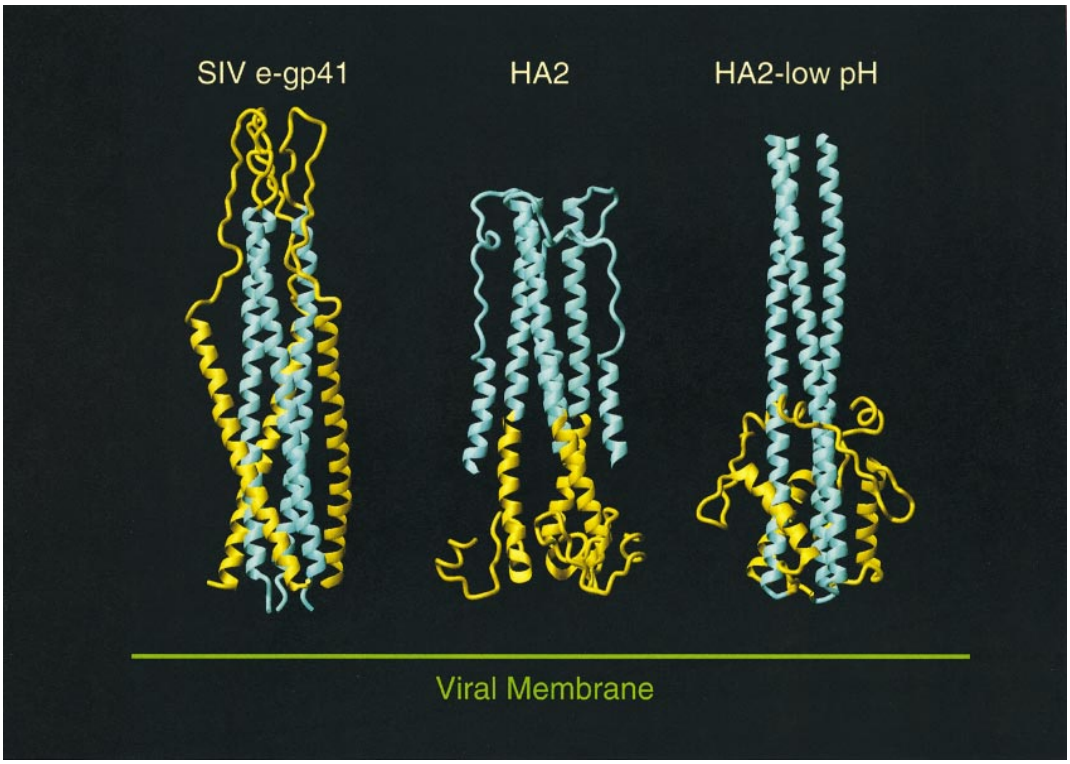

Fig. 8. Comparison of the structure of SIV e-gp41 with the neutral (Wilson *et al*., 1981; Weis *et al*., 1990) and low (Bullough *et al*., 1994) pH forms of HA2. N-terminal helices are colored cyan and the C-terminal region is colored yellow. Note that the N-terminal helices in hemagglutinin are located on the exterior of the molecule, whereas in e-gp41 they are located in the interior.

monoclonal antibody KK20, that is located in the loop and comprises residues 77–99 (Sattentau *et al*., 1993). The explanation for the observation that gp120 association appears to be minimally, if at all, perturbed in the *in vivo* assays by mutations that decrease the stability of the trimer is probably due to a counter effect arising from gp120 binding itself that shifts the monomer–trimer equilibrium in favor of the trimeric form of e-gp41.

Previous mutagenesis studies of HIV-1 gp120 have suggested that residues 4–14 of the C1 subdomain are involved in binding e-gp41 (Helseth *et al*., 1991; Ivey-Hoyle *et al*., 1991). We note that this region is very hydrophobic in both HIV-1 and SIV gp120 (LWVTVYYGVPV and QYVTVFYGVPA, respectively; Douglas *et al*., 1997). Moreover, HIV-1 gp120 exhibits a five residue insertion (residues 34–38) within the C1 subdomain with respect to SIV gp120. Interestingly, this insertion may correlate with a four residue insertion (located between residues 99 and 100 of SIV e-gp41) in the loop region of HIV-1 e-gp41 relative to SIV e-gp41 (Figure 1), lending further circumstantial support for the binding of the gp120 C1 region to the e-gp41 loop region.

Relationship of e-gp41 to influenza virus hemagglutinin

gp41 (Freed and Martin, 1995) and influenza virus HA2 (Skehel *et al*., 1996) share a number of common features: both proteins are formed by proteolytic cleavage of a large precursor and mediate virus–target membrane fusion; in both proteins, the fusion peptide is located at the N-terminus and the transmembrane domain at the C-terminus of the ectodomain; and both proteins display a trimeric coiled-coil arrangement of the N-terminal helices. There are, however, a number of important distinctions, both functional and structural, between HA2 and gp41 which suggest that in the case of e-gp41 there is no need to invoke the large spring-loaded (Skehel *et al*., 1982; Carr and Kim, 1993) conformational change that occurs upon the conversion of the nonfusogenic, neutral pH form of HA2 (Figure 8, middle panel; Wilson *et al*., 1981; Weis *et al*., 1990) to the fusogenic, low pH form of HA2 (Figure 8, right panel; Bullough *et al*., 1994). First, from a purely functional viewpoint, gp41- and HA2-mediated fusion involve a different series of events. In particular, HA2 is internalized by endocytosis and mediates fusion of the viral and endosome membranes while gp41 mediates fusion of the viral and outer membranes; in addition, HA2-mediated fusion is triggered by low pH (Wiley and Skehel, 1987) while gp41-mediated fusion is pH-independent (Stein *et al*., 1987). Secondly, the loop region that connects the first and second helices in the neutral form of HA2 does not contain any prolines, comprises a sequence with a high degree of helix propensity and, as demonstrated by X-ray crystallography, undergoes a transition to a helical coiled-coil at low pH (Bullough *et al*., 1994). In contrast, the loop connecting the N- and C-terminal helices in SIV e-gp41 contains three prolines (two prolines in HIV-1 e-gp41) and exhibits no propensity to form a helical coiled-coil, as ascertained using the program MultiCoils (Wolf *et al*., 1997). Thirdly, as is evident from Figure 8, e-gp41 is topologically very different from either the neutral or low pH forms of HA2. In HA2, the N-terminal helix is always located on the outside of the molecule. In the neutral form the N-terminal helix is connected by a loop to a second helix which lies internal to it. In the low pH form, the loop is converted to a helix

such that a single contiguous helix in a trimer coiled-coil arrangement lying on the outside of the molecule is formed. Thus, there is no topological impediment to the conformational transition from the neutral to the low pH form of HA2 even though this involves a movement of ~100 Å in the position of the N-terminus. In contrast, the N-terminal helix of the three subunits of e-gp41 forms a trimeric coiled-coil within the interior of the protein, fully surrounded by the C-terminal helices on the exterior. A conformational transition of the type observed in HA2 would therefore require the trimeric coiled-coil formed by the N-terminal helix of e-gp41 to be everted, which seems unlikely. Such an event would presumably require prior dissolution of the trimeric coiled-coil as well as the complete dissociation of the intra- and intermolecular contacts between the N- and C-terminal helices. Consequently, the significant structural, topological and functional differences between HA2 and e-gp41 suggest that parallels between the structure and function of the two systems, although intriguing, may not necessarily exist.

Inhibition of gp41-mediated fusion by peptides

Peptides derived from both the N-terminal (N-peptide) and C-terminal (C-peptide) helices of gp41 have been shown to inhibit fusion in a dominant-negative manner (Wild *et al*., 1992, 1994) and it has been suggested that these data provide evidence for a large conformational change from the nonfusogenic to the fusogenic states (Furuta *et al*., 1998; Munoz-Barroso *et al*., 1998). While the e-gp41 trimer is highly stable with a T_m in excess of 100°C, analytical ultracentrifugation has shown that it exists as a mixture of monomer and trimer with a self-association constant of ~1.5×10^{11} M^{-2} for the SIV e-gp41 construct employed here and ~4.5×10^{11} M^{-2} for the equivalent HIV-1 e-gp41 construct (Wingfield *et al*., 1997). Thus, for gp41 trimer concentrations of 500, 50, 5 and 0.5 μM, the concentration of coexisting monomeric gp41 will be 15, 7, 3.2 and 1.5 μM, respectively, in the case of SIV e-gp41, and 10, 4.7, 2.2 and 1 μM, respectively, in the case of HIV-1 e-gp41. (Note that under the conditions employed in the present NMR study where the total concentration of SIV e-gp41 is ~2.5 mM in monomer units, <1% of the total e-gp41 will be present in the monomeric form.)

A possible mechanism of fusion inhibition by the peptides may therefore simply involve the scenario outlined in the scheme shown in Figure 9A. gp41 exists in an equilibrium between monomer and trimer. In the presence of excess inhibitory peptide, the equilibrium is driven from homotrimeric gp41 to a heterotrimer of gp41 and N- or C-peptide. Since the peptides are only effective upon gp120 dissociation, it follows that the absence of fusogenic activity displayed by the gp41–peptide heterotrimers is due to the fact that the heterotrimers can no longer present a sufficient number of fusion peptides to the target membrane for effective fusion to take place. The data also suggest that the trimeric state is stabilized upon gp120 binding such that heterotrimers (which would be expected to bind less tightly to gp120 since they do not possess a trimeric loop structure) cannot be formed in the presence of bound gp120 (Furuta *et al*., 1998). It has also been shown that the N-peptide is approximately three orders of magnitude less active than the C-peptide. This is

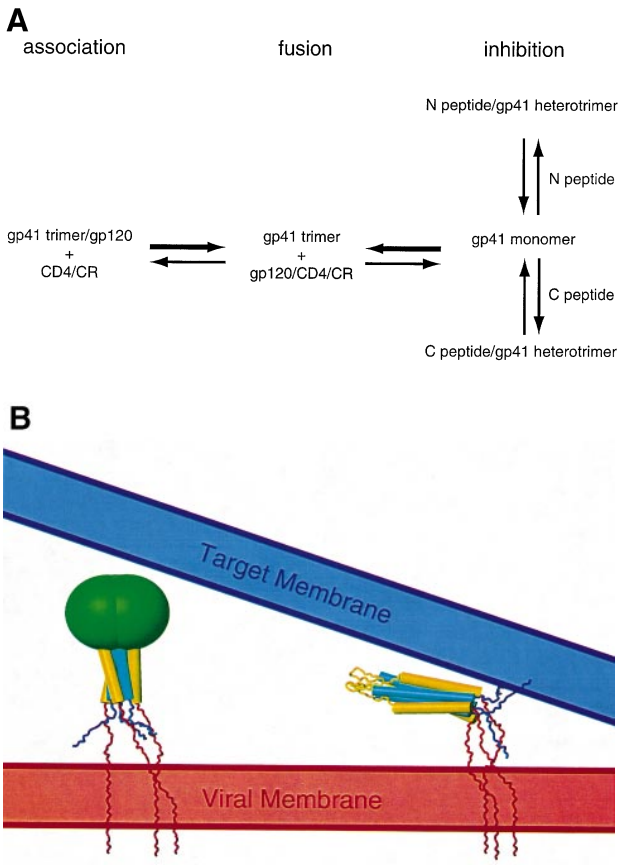

Fig. 9. Model for (**A**) the inhibitory effects of peptides derived from the N- and C-terminal helices and (**B**) gp41-mediated fusion. In (B) residues 1–26, which include the fusion peptide (residues 1–15) and which lie N-terminal to the e-gp41 construct employed in the present study (residues 27–149), are blue; the residues C-terminal to e-gp41 which comprise a 20 residue linker, followed by the transmembrane segment are red; the N-terminal region of SIV e-gp41 (residues 27–80) is cyan, and the C-terminal region (residues 81–149) yellow; the helices in gp41 are displayed as cylinders; gp120 is shown as a green sphere.

not surprising, since the N-peptide itself forms oligomers, whereas the C-peptide is monomeric (Blacklow *et al*., 1995; Lu *et al*., 1995). Thus, in the case of the N-peptide, the effective concentration of monomeric N-peptide available to interact with monomeric gp41 to form heterotrimers will be much reduced. Moreover, C-peptide-mediated inhibition significantly decreases upon the addition of stoichiometric amounts of N-peptide (Lu *et al*., 1995), consistent with the notion that the presence of monomeric C- or N-peptide is essential for inhibition.

Model for gp41-mediated membrane fusion

The structure of SIV e-gp41 presented here combined with the mutational and peptide data suggest to us the following mechanism for gp41-mediated fusion (Figure 9B). The C-terminus of e-gp41 is tethered via a flexible 20 residue linker to the transmembrane segment which anchors gp41 to the viral membrane (the transmembrane domain of SIV gp41 starts at residue 168). As a result, the orientation of the ectodomain of gp41 with regard to the plane of the viral membrane is likely to be highly variable and dynamic. For fusion to take place, one would predict that the ectodomain should lie approximately parallel to the viral membrane, in order to permit the

N-terminal fusion peptides to have access to the target membrane (Figure 9B). Because of the large size of gp120 relative to gp41, such an orientation is not accessible to the gp41–gp120 complex. The same steric hindrance mechanism would apply to the complex formed between gp41 and the neutralizing monoclonal antibody 2F5 (Muster *et al*., 1993) since the latter binds to the base of e-gp41 (Figure 4c). Upon dissociation of gp120 from gp41 subsequent to the interaction of gp120 to the CD4 and chemokine receptors, the full range of orientations of e-gp41 with regard to the viral membrane becomes accessible, permitting the N-terminal fusion peptides to contact and insert transitorily into the target membrane as a result of random Brownian motion.

This model has several attractive features. First, it is relatively simple since it does not require a large, topologically complex, conformational change that has not been observed to date by any structural or physicochemical technique. Indeed, only a single conformation of e-gp41 has been observed by NMR (this paper; Caffrey *et al*., 1997), circular dichroism (Wingfield *et al*., 1997), EPR (Rabenstein and Shin, 1996), electron microscopy (Weissenhorn *et al*., 1996) and X-ray crystallography (Chan *et al*., 1997; Tan *et al*., 1997; Weissenhorn *et al*., 1997). Secondly, the fusion reaction is relatively nonspecific, in agreement with the tolerance for mutations in the fusion peptide (Steffy *et al*., 1992). Thirdly, the loop region, which is immunodominant, is exposed upon gp120 dissociation (Xu *et al*., 1991). Finally, we note that this model is not dissimilar from that proposed for the SNARE proteins, which mediate cellular membrane fusion by forming a complex between discrete proteins located on different membranes in a manner analogous to the association between the N- and C-terminal helices of e-gp41 (Weber *et al*., 1998).

Materials and methods

SIV e-gp41 (residues 27–149) with Cys86 and Cys92 mutated to alanine was expressed and purified as described previously (Caffrey *et al*., 1997). Unlabeled (i.e. at natural isotopic abundance), and uniformly ^{15}N, ^{15}N/^{13}C-, ^{15}N/^{2}H-, ^{15}N/^{13}C/^{2}H-labeled samples were prepared by growing the bacteria in either H_2O or $^{2}H_2O$ (for ^{2}H labeling) in minimal medium using $^{15}NH_4Cl$ and $^{13}C_6$-glucose as sole nitrogen and carbon sources, respectively. Labeling with ^{15}N and ^{13}C was >95%, and labeling with ^{2}H was >80%. In addition, stereospecific assignments of valine and leucine methyl groups were obtained using a 10% ^{13}C-labeled sample (Johnson *et al*., 1996). Samples used to detect intermolecular NOEs comprised 1:1 mixtures of ^{13}C/^{15}N/^{1}H:^{12}C/^{14}N/^{1}H-, ^{13}C/^{14}N/^{1}H:^{12}C/^{15}N/^{2}H- and ^{13}C/^{15}N/^{2}H:^{12}C/^{14}N/^{1}H-labeled SIV e-gp41. Samples for NMR contained ~2.5 mM SIV e-gp41 (monomer concentration) in 50 mM deuterated sodium formate pH 3.0, and all spectra were recorded at 45°C on Bruker DMX500, DMX600 and DMX750 spectrometers. All spectra were processed using NmrPipe (Delaglio *et al*., 1995) and analyzed using the programs PIPP, CAPP and STAPP (Garrett *et al*., 1991).

A description of the procedures and experiments used to obtain the ^{1}H, ^{15}N and ^{13}C assignments was provided in Caffrey *et al*. (1997). Interproton distance restraints were derived from multidimensional NOE spectra (Clore and Gronenborn, 1991, 1998) with mixing times ranging from 60 to 120 ms. Experiments included 3D ^{13}C-separated, ^{15}N-separated, ^{13}C-separated/^{12}C-filtered, ^{13}C-separated/^{15}N-filtered and ^{15}N-separated/^{13}C-filtered NOE spectra, a 3D ^{15}N-separated ROE spectrum, and 4D ^{13}C/^{15}N-separated, ^{15}N/^{15}N-separated and ^{13}C/^{13}C-separated NOE spectra. The NOEs used for distance restraints were classified as: strong (1.8–2.7 or 1.8–2.9 Å for NOE of NH), medium (1.8–3.3 or 1.8–3.5 Å for NOE of NH), weak (1.8–5.0 Å) or very weak (1.8–6.0 Å). For distances involving methyl protons, 0.5 Å was added to account for the higher apparent intensity of methyl protons. In the case of nonstereospecifically assigned protons and intersubunit NOEs, distances were represented by a $\Sigma(r^{-6})^{-1/6}$ sum (Nilges, 1993). Hydrogen bonding restraints (two per hydrogen bond where $r_{NH\text{-}O}$ = 1.5–2.8 Å and $r_{N\text{-}O}$ = 2.4–3.5 Å) were deduced from NH exchange experiments, backbone NOEs and backbone chemical shifts (Caffrey *et al*., 1997). ϕ and ψ torsion angle restraints were derived from $^{3}J_{HN\alpha}$ (Bax *et al*., 1994) and $^{3}J_{C'C'}$ couplings (Hu and Bax, 1996), the three-bond amide deuterium isotope effect on $^{13}C_\alpha$ chemical shifts (Ottiger and Bax, 1997) and a database analysis of backbone (N, HN, C_α, C_β, C′, H_α) chemical shifts using the program TALOS (Cornilescu *et al*., 1998). χ_1 and χ_2 torsion angle restraints were derived from analysis of heteronuclear $^{3}J_{CC}$, $^{3}J_{NC\gamma}$ and $^{3}J_{COC\gamma}$ couplings (Bax *et al*., 1994; Hu and Bax, 1997; Hu *et al*., 1997) and ROE and short-mixing-time NOE experiments. Minimum error ranges employed for ϕ, ψ, χ_1 and χ_2 were ±30, ±45, ±20 and ±30°, respectively. Structures were calculated by simulated annealing in torsion angle space (Stein *et al*., 1997) starting from three extended strands, followed by conventional simulated annealing in cartesian coordinate space (Nilges *et al*., 1988), using the program CNS (Brünger *et al*., 1998), which was adapted to incorporate pseudopotentials for $^{3}J_{HN\alpha}$ and $^{3}J_{C'C'}$ coupling contants (Garret *et al*., 1994), three-bond amide deuterium isotope effects on $^{13}C_\alpha$ shifts (Garrett *et al*., 1994), secondary $^{13}C_\alpha$ and $^{13}C_\beta$ chemical shifts (Kuszweski *et al*., 1995), and a conformational database (Kuszewski *et al*., 1996, 1997). A pseudopotential term for noncrystallographic symmetry was employed in all calculations. Figures were generated using the programs MOLMOL (Koradi *et al*., 1996) and GRASP (Nicholls *et al*., 1991). Electrostatic calculations were performed with GRASP (Nicholls *et al*., 1991). Calculation, ranking and mapping of surface hydrophobic clusters was carried out as previously described (Covell *et al*., 1994; Young *et al*., 1994). The coordinates for the final 40 simulated annealing structures, together with the coordinates for the corresponding restrained regularized mean structure and a complete list of experimental NMR restraints have been deposited in the Brookhaven Protein Data Bank (Accession codes 2EZO, 2EZP, 2EZOMR).

Acknowledgements

We thank Dan Garrett, Frank Delaglio and Gabriel Cornilescu for software support, Rolf Tschudin for technical support, and Ad Bax for useful discussions. This work was supported by the AIDS Targeted Antiviral Program of the Office of the Director of the National Institutes of Health (to G.M.C., A.M.G. and the Protein Expression Laboratory of NIAMSD).

References

Bax,A. and Grzesiek,S. (1993) Methodological advances in protein NMR. *Acct. Chem. Res.*, **26**, 131–138.

Bax,A., Vuister,G.W., Grzesiek,S., Delaglio,F., Wang,A.C., Tschudin,R. and Zhu,G. (1994) Measurement of homo- and heteronuclear J couplings from quantitative J correlation. *Methods Enzymol.*, **239**, 79–125.

Blacklow,S., Lu,M. and Kim,P. (1995) A trimeric subdomain of the simian immunodeficiency virus envelope glycoprotein. *Biochemistry*, **34**, 14955–14962.

Brünger,A.T. *et al.* (1998) Crystallography and NMR system (CNS): a new software suite for macromolecular structure determination. *Acta Crystallogr. D*, in press.

Bullough,P.A., Hughson,F.M., Skehel,J.J. and Wiley,D.C. (1994) Structure of influenza haemagglutinin at the pH of membrane fusion. *Nature*, **371**, 37–43.

Caffrey,M., Cai,M., Kaufman,J., Stahl,S.J., Wingfield,P.T., Gronenborn, A.M. and Clore,G.M. (1997) Determination of the secondary structure and global topology of the 44 kDa ectodomain of gp41 of the simian immunodeficiency virus by multidimensional nuclear magnetic resonance spectroscopy. *J. Mol. Biol.*, **271**, 819–826.

Cao,J., Bergeron,L., Helseth,E., Thali,M., Repke,H. and Sodroski,J. (1993) Effects of amino acid changes in the extracellular domain of the human immunodeficiency virus type 1 gp41 envelope glycoprotein. *J. Virol.*, **67**, 2747–2755.

Carr,C.M. and Kim,P.S. (1993) A spring-loaded mechanism for the conformational change of influenza hemagglutinin. *Cell*, **73**, 823–832.

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

Chen,S.S.-L. (1994) Functional role of the zipper motif region of human immunodeficiency virus type 1 transmembrane protein gp41. *J. Virol.*, **68**, 2002–2010.

Clore,G.M. and Gronenborn,A.M. (1991) Structures of larger proteins in solution: three- and four-dimensional heteronuclear NMR spectroscopy. *Science*, **252**, 1390–1399.

Clore,G.M. and Gronenborn,A.M. (1998) Determining the structures of large proteins and protein complexes by NMR. *Trends Biotechnol.*, **16**, 22–34.

Coffin,J.M. (1986) Genetic variation in AIDS viruses. *Cell*, **46**, 1–4.

Cornilescu,G., Delaglio D. and Bax,A. (1998) TALOS: a program for deriving protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR*, in press.

Covell,D.G., Smythers,G.W., Gronenborn,A.M. and Clore,G.M. (1994) Analysis of hydrophobicity in the α and β chemokine families and its relevance to dimerization. *Protein Sci.*, **3**, 2064–2072.

Dedra,D., Gu,R. and Tatner,L. (1992) Role of asparagine-linked glycosylation in human immunodeficiency virus type I transmembrane envelope function. *Virology*, **187**, 377–382.

Delaglio,F., Grzesiek,S., Vuister,G.W., Zhu,G., Pfeifer,J. and Bax,A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR*, **6**, 277–293.

Douglas,N.W., Munro,G.H. and Daniels,R.S. (1997) HIV/SIV glycoproteins: structure–function relationships. *J. Mol. Biol.*, **273**, 122–149.

Freed,E.O. and Martin,M.A. (1995) The role of human immunodeficiency virus type 1 envelope glycoproteins in virus infection. *J. Biol. Chem.*, **270**, 23883–23886.

Furuta,R.A., Wild,C.T., Weng,Y.K. and Weiss,C.D. (1998) Capture of an early fusion-active conformation of HIV-1 gp41. *Nature Struct. Biol.*, **5**, 276–279.

Gallaher,W. (1987) Detection of a fusion peptide sequence in the transmembrane proteins of human immunodeficiency virus. *Cell*, **50**, 327–328.

Garrett,D., Powers,R., Gronenborn,A. and Clore,G. (1991) A common sense approach to peak picking in two-, three- and four-dimensional spectra using automatic computer analysis of contour diagrams. *J. Magn. Reson.*, **94**, 214–220.

Garrett,D.S., Kuszewski,J., Hancock,T.J., Lodi,P.J., Vuister,G.W., Gronenborn,A.M. and Clore,G.M. (1994) The impact of direct refinement against three bond HN-$C^{\alpha}H$ coupling constants on protein structure determination by NMR. *J. Magn. Reson.*, **104**, 99–103.

Ivey-Hoyle,M., Clark,R.K. and Rosenberg,M. (1991) The NH2-terminal 31 amino acids of human immunodeficiency virus type-1 envelope protein gp120 contains a potential gp41 contact site. *J. Virol.*, **65**, 2682–2685.

Helseth,E., Olshefsky,U., Furman,C. and Sodrosky,J. (1991) Human immunodeficiency virus type-1 gp120 envelope glycoprotein regions important for association with the gp41 transmembrane glycoprotein. *J. Virol.*, **65**, 2119–2123.

Hu,J.S. and Bax,A. (1996) Measurement of three-bond ^{13}C-^{13}C J couplings between carbonyl and carbonyl/carboxyl carbons in isotopically enriched proteins. *J. Am. Chem. Soc.*, **118**, 8170–8171.

Hu,J.S. and Bax,A. (1997) χ_1 angle information from a simple two-dimensional NMR experiment that identifies *trans* $^3J_{NC\gamma}$ coupling in isotopically enriched proteins. *J. Biomol. NMR*, **9**, 323–328.

Hu,J.S., Grzesiek,S. and Bax,A. (1997) Two dimensional NMR methods for determining χ_1 angles of aromatic residues in proteins from three-bond $J_{C'C\gamma}$ and $J_{NC\gamma}$ couplings. *J. Am. Chem. Soc.*, **119**, 1803–1804.

Johnson,P.E., Joshi,M.D., Tomme,P., Kilburn,D.G. and McIntosh,L.P. (1996) Structure of the N-terminal cellulose-binding domain of *Cellulomonas fimi* CenC determined by nuclear magnetic resonance spectroscopy. *Biochemistry*, **35**, 14381–14394.

Jones,S. and Thornton,J.M. (1996) Principles of protein–protein interactions. *Proc. Natl Acad. Sci. USA*, **93**, 13–20.

Koradi,R., Billeter,M. and Wuthrich,K. (1996) MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.*, **14**, 52–55.

Kuszewski,J., Gronenborn,A.M. and Clore,G.M. (1995) The impact of direct refinement against $^{13}C_{\alpha}$ and $^{13}C_{\beta}$ chemical shifts on protein structure determination by NMR. *J. Magn. Reson. Ser. B*, **106**, 92–96.

Kuszewski,J., Gronenborn,A.M. and Clore,G.M. (1996) Improving the quality of NMR and crystallographic protein structures by means of a conformational database potential derived from structure databases. *Protein Sci.*, **5**, 1067–1080.

Kuszewski,J., Gronenborn,A.M. and Clore,G.M. (1997) Improvements and extensions in the conformational database potential for the refinement of NMR and X-ray structures of proteins and nucleic acids. *J. Magn. Reson.*, **125**, 171–177.

Laskowski,R.A., MacAruther,M.W., Moss,D.S. and Thornton,J.M. (1993) PROCHECK: a program to check stereochemical quality of protein structures. *J. Appl. Crystallogr.*, **26**, 283–291.

Lu,M., Blacklow,S. and Kim,P. (1995) A trimeric structural domain of the HIV-1 transmembrane glycoprotein. *Nature Struct. Biol.*, **2**, 1075–1082.

Moore,J.P., Jameson,B.A., Weiss,R.A. and Sattentau,Q.J. (1993) The HIV-cell fusion reaction. In Betz,J. (ed.), *Viral Fusion Mechanisms*. CRC Press, Boca Raton, FL, pp. 233–289.

Moore,J.P., Trkola,A. and Dragic,T. (1997) Co-receptors for HIV-1 entry. *Curr. Opin. Immunol.*, **9**, 551–562.

Munoz-Barroso,I., Durell,S., Sakaguchi,K., Apella,E. and Blumenthal,R. (1998) Dilation of the human immunodeficiency virus-1 envelope glycoptorein fusion pore revealed by the inhibitory action of a synthetic peptide from gp41. *J. Cell Biol.*, **140**, 315–323.

Muster,T., Steindl,F., Purtscher,M., Trkola,A., Klima,A., Himmler,G., Ruker,F. and Katinger,H. (1993) A conserved neutralizing epitope on gp41 of human immunodeficiency virus type I. *J. Virol.*, **67**, 6642–6647.

Nicholls,A., Sharp,K.A. and Honig,B. (1991) Proteins folding and association: insights from the interfacial and thermodynamic porperties of hydrocarbons. *Proteins Struct. Funct. Genet.*, **11**, 281–296.

Nilges,M. (1993) A calculational strategy for the structure determination of symmetric dimers by ^1H NMR. *Proteins Struct. Funct. Genet.*, **17**, 297–309.

Nilges,M., Clore,G.M. and Gronenborn,A.M. (1988) Determination of three-dimensional structures of proteins from interproton distances data by hybrid distance geometry-dynamical simulated annealing calculations. *FEBS Lett.*, **229**, 317–324.

Ottiger,M. and Bax,A. (1997) An empirical correlation between amide deuterium isotope shift effects on $^{13}C_{\alpha}$ chemical shifts and protein backbone confirmation. *J. Am. Chem. Soc.*, **119**, 8070–8075.

Rabenstein,M.D. and Shin,Y.-K. (1996) HIV-1 gp41 tertiary structure studied by EPR spectroscopy. *Biochemistry*, **35**, 13922–13928.

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

Sattentau,Q.J, Zolla-Pazner,S. and Poignard,P. (1995) Epitope exposure on functional oligomeric HIV-1 gp41 molecules. *Virology*, **206**, 713–717.

Schulz,T.F., Jameson,B.A., Lopalco,L., Sicardi,A.G., Weiss,R.A. and Moore,J.P. (1992) Conserved structural features in the interaction between retroviral surface and transmembrane glycoproteins. *AIDS Res. Hum. Retroviruses*, **8**, 1571–1580.

Skehel,J.J., Bayley,P.M., Brown,E.B., Martin,S.R., Waterfield,M.D., White,J.M., Wilson,I.A. and Wiley,D.C. (1982) Changes in the conformation of influenza virus hemagluttinin at the pH optimum of virus-mediated membrane fusion. *Proc. Natl Acad. Sci. USA*, **79**, 968–972.

Skehel,J.J., Bizebard,T., Bullough,P.A., Hughson,F.M., Knossow,M., Steinbauer,D.A., Wharton,S.A. and Wiley,D.C. (1996) Membrane fusion by influenza hemagluttinin. *Cold Spring Harbor Symp. Quant. Biol.*, **60**, 573–580.

Steffy,K.R., Kraus,G., Looney,D.J. and Wong-Staal,F. (1992) Role of the fusogenic peptide sequence in syncytium induction and infectivity of human immunodeficiency virus type 2. *J. Virol.*, **66**, 4532–4535.

Stein,B.S., Gowda,S.D., Lifson,J.D., Penhallow,R.C., Bensch,K.G. and Engleman,E.G. (1987) pH-independent HIV entry into CD4-positive T cells via virus envelope fusion to the plasma membrane. *Cell*, **49**, 659–668.

Stein,E.G., Rice,L.M. and Brunger,A.T. (1997) Torsion-angle molecular dynamics as a new efficient tool for NMR structure calculation. *J. Magn. Reson.*, **124**, 154–164.

Tan,K., Liu,J.-H., Wang,J.-H., Shen,S. and Lu,M. (1997) Atomic structure of a thermostable subdomain of HIV-1 gp41. *Proc. Natl Acad. Sci. USA*, **94**, 12303–12308.

Weber,T., Zemelman,B.V., McNew,J.A., Westermann,B., Gmachi,M., Parlati,F., Sollner,T.H. and Rothman,J.E. (1998) SNAREpins: minimal machinery for membrane fusion. *Cell*, **92**, 759–772.

Weis,W., Brünger,A.T., Skehel,J.J. and Wiley,D.C. (1990) Refinement of the influenza virus hemagglutinin by simulated annealing. *J. Mol. Biol.*, **212**, 737–761.

Weissenhorn,W., Wharton,S., Calder,L., Earl,P., Moss,B., Aliprandis,E., Skehel,J.J. and Wiley,D. (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**, 1507–1514.

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

Wild,C.T., Oas,T., McDanal,C.B., Bolognesi,D. and Matthews,T.J. (1992) A synthetic peptide inhibitor of human immunodeficiency virus replication: correlation between solution structure and viral inhibition. *Proc. Natl Acad. Sci. USA*, **89**, 10537–10541.

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

Wiley,D.C. and Skehel,J.J. (1987) The structure and function of the hemagglutinin membrane glycoprotein of influenza virus. *Annu. Rev. Biochem.*, **56**, 365–394.

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

Wingfield,P.T., Stahl,S.J., Kaufman,J., Zlotnick,A., Hyde,C.C., Gronenborn,A.M. and Clore,G.M. (1997) The extracellular domain of immunodeficiency virus gp41 protein. Expression in *Escherichia coli*, purification and crystallization. *Protein Sci.*, **6**, 1653–1660.

Wolf,E., Kim,P.S. and Berger,B. (1997) MultiCoil: a program for predicting two- and three-stranded coiled coils. *Protein Sci.*, **6**, 1179–1189.

Xu,J.-X., Gorny,M., Palker,T., Karwowska,S. and Zolla-Pasner,S. (1991) Epitope mapping of two immunodominant domains of gp41, the transmembrane protein of human immunodeficiency virus type 1, using ten human monoclonal antibodies. *J. Virol.*, **65**, 4832–4838.

Young,L., Jernigan,R.C. and Covell,D.G. (1994) A role for surface hydrophobicity in protein–protein recognition. *Protein Sci.*, **3**, 717–729.

Received May 13, 1998; revised June 25, 1998;
accepted June 26, 1998