

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

Michael Caffrey¹, Mengli Cai¹, Joshua Kaufman², Stephen J. Stahl²
Paul T. Wingfield², Angela M. Gronenborn^{1*} and G. Marius Clore^{1*}

¹Laboratory of Chemical Physics, Building 5, National Institute of Diabetes and Digestive and Kidney Diseases National Institutes of Health Bethesda, MD 20892-0520 USA

²Protein Expression Laboratory Building 6B, National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health, Bethesda MD 20892-2775, USA

The gp41 protein of the human (HIV) and simian (SIV) immunodeficiency viruses is part of the envelope glycoprotein complex gp41/gp120 which plays an essential role in viral infection. We present a multidimensional NMR study on the trimeric 44 kDa soluble ectodomain of SIV gp41 (e-gp41), comprising residues 27 to 149. Despite the large molecular weight and very limited spectral dispersion, complete backbone ¹H, ¹³C, ¹³CO and ¹⁵N assignments have been made using a combination of triple resonance experiments on uniformly ¹³C/¹⁵N and ²H/¹³C/¹⁵N-labeled samples. The secondary structure of SIV e-gp41, derived on the basis of ¹³C chemical shifts, NH exchange rates, medium range nuclear Overhauser enhancements (NOE), and ³J_{HNα} coupling constants, consists of a 49 residue helix at the N terminus (residues 29 to 77) and a 40 residue helix at the C terminus (residues 108 to 147), connected by a 30 residue loop which does not display any of the characteristics of regular secondary structure. The cross-peak intensities of the loop region in scalar correlation experiments suggests that it is more mobile than the core helical regions. The presence, however, of numerous long range NOEs, both intra and inter-subunit, within the loop indicates that it adopts a well-defined structure in which the loops from the three subunits interact with each other. Based on a number of long range intra and inter-subunit NOEs, a topological model is presented for the symmetric SIV e-gp41 trimer in which the N-terminal helices are packed within the protein interior in a parallel trimeric coiled-coil arrangement, while the C-terminal helices are located on the protein exterior, oriented antiparallel to the N-terminal helices.

© 1997 Academic Press Limited

*Corresponding authors

Keywords: gp41; SIV; HIV; NMR; global topology

The gp41 protein of the human (HIV) and simian (SIV) immunodeficiency viruses is part of the gp41/gp120 envelope glycoprotein complex. gp41 and gp120 are cleavage products of gp160, which is encoded by the viral *env* gene (Allan *et al.*, 1985; Veronese *et al.*, 1985). During the initial stages of infection, the gp41/gp120 complex associates with the CD4 receptor (Klatzmann *et al.*, 1984; Dalglish *et al.*, 1984), as well as with the newly discovered chemokine coreceptors (Feng *et al.*, 1996; Deng

et al., 1996; Litwin *et al.*, 1996). Subsequently, gp41 dissociates from gp120, associates with the target membrane, and mediates fusion of the viral and cellular membranes by a process that involves the N-terminal hydrophobic region of gp41, termed the fusion peptide (Gallaher, 1987). gp41 consists of four distinct functional regions: a fusion peptide, an ectodomain, a transmembrane domain, and a cytoplasmic domain. Biochemical studies have shown that the ectodomain of SIV and HIV gp41 (e-gp41) has high helical content and exists in solution as a multimer (Blacklow *et al.*, 1995; Lawless *et al.*, 1996; Lu *et al.*, 1995; Weissenhorn *et al.*, 1996). Protease protection experiments combined with circular dichroism measurements on

Abbreviations used: HIV, human immunodeficiency virus; SIV, simian immunodeficiency virus; e-gp41, ectodomain of gp41; HSQC, heteronuclear single quantum coherence; NOE, nuclear Overhauser effect.

HIV and SIV e-gp41 suggest the presence of N and C-terminal helices oriented in an antiparallel manner (Blacklow *et al.*, 1995; Lu *et al.*, 1995). This is consistent with the results from a recent EPR study on spin-labeled synthetic peptides corresponding to parts of the putative N and C-terminal helices of HIV e-gp41 (Rabenstein & Shin, 1996).

The absence of high resolution structural information makes HIV or SIV gp41 an attractive candidate for study by NMR, although its poor solubility and high molecular weight render such studies technically challenging. We have recently overcome some of these difficulties by preparing the ectodomain of SIV gp41, consisting of residues 27 to 149 (Wingfield *et al.*, 1997). This construct lacks the fusion peptide necessary for membrane fusion and, since it is expressed in *Escherichia coli*,

is not glycosylated. Mutagenesis studies, however, have demonstrated that glycosylation of HIV gp41 does not play a critical role in folding, processing or function (Dedra *et al.*, 1992). The present SIV e-gp41 construct contains a highly conserved region that is thought to associate with gp120 (Kowalski *et al.*, 1987), as well as the binding site for a monoclonal antibody that blocks viral infection *in vitro* (Muster *et al.*, 1993). At acidic pH values (<4), SIV e-gp41 forms a soluble 44 kDa trimeric complex (Wingfield *et al.*, 1997). (Note that at pH values above 4.5, SIV e-gp41 precipitates out of solution.) The solubility properties of SIV e-gp41 and the ability to express and label it with stable NMR active isotopes has made e-gp41 amenable to heteronuclear NMR studies of its structure and dynamics. Here, we present the complete backbone

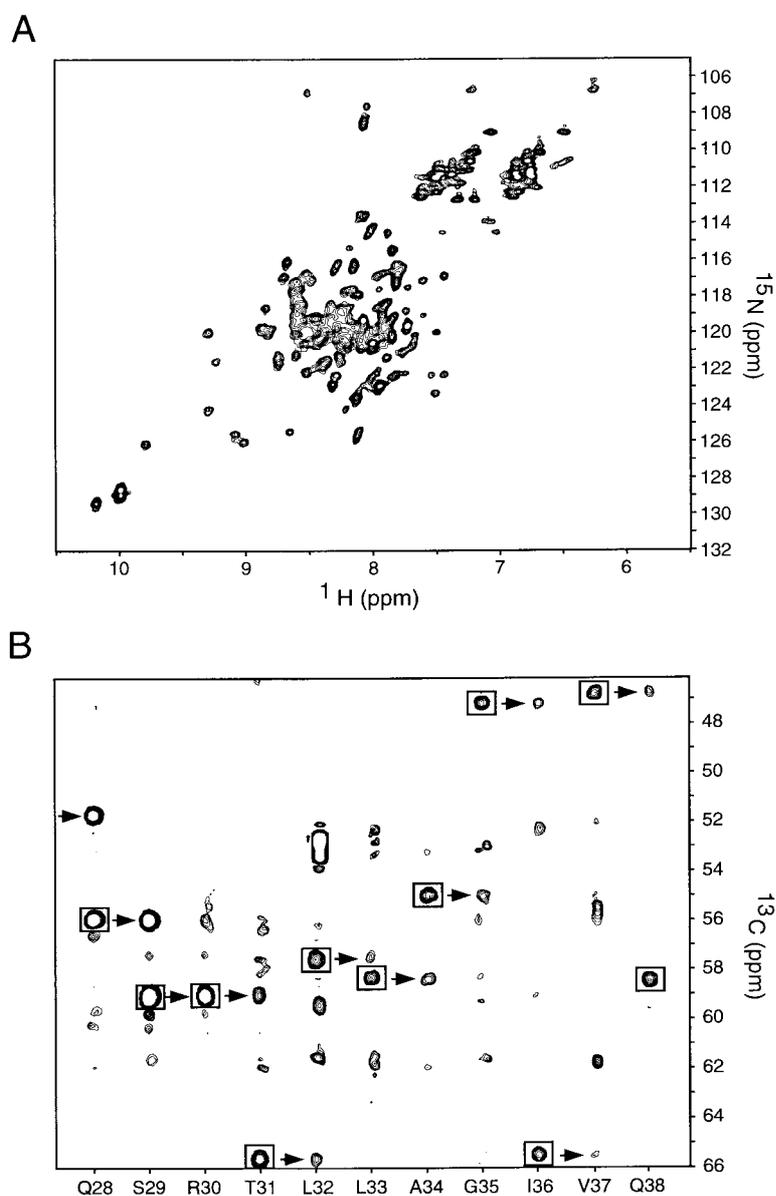


Figure 1. A, 2D ^1H - ^{15}N HSQC spectrum of perdeuterated ^{15}N -labeled SIV e-gp41 at 45°C . B, Strip taken from the 3D HNCA experiment recorded at 45°C on $^2\text{H}/^{13}\text{C}/^{15}\text{N}$ -labeled SIV e-gp41 illustrating through-bond sequential assignments for residues 28 to 38. The intraresidue $\text{C}^\alpha(i)\text{-N}(i)\text{-HN}(i)$ correlations are boxed, while the interresidue $\text{C}^\alpha(i-1)\text{-N}(i)\text{-HN}(i)$ correlations are denoted by arrows. The signals for Val37 are folded in the ^{13}C dimension. Genetic constructs, protein expression and purification were as described by Wingfield *et al.* (1997). Uniform (>95%) ^{15}N and ^{13}C labeling was carried out by growing the bacteria in minimal medium using $^{15}\text{NH}_4\text{Cl}$ and $^{13}\text{C}_6$ -glucose as the sole nitrogen and carbon sources, respectively. Perdeuteration (>80% ^2H labeling) was achieved by growth in $^2\text{H}_2\text{O}$ instead of H_2O , as described by Garrett *et al.* (1997). For the experiments in H_2O , the sample conditions were 2.5 mM e-gp41 (monomer concentration) in 50 mM deuterated sodium formate (pH 3.0) and 90% $\text{H}_2\text{O}/10\%$ $^2\text{H}_2\text{O}$. For the experiments in $^2\text{H}_2\text{O}$, a solution containing 2.5 mM e-gp41 (monomer concentration) and 50 mM deuterated sodium formate in H_2O was lyophilized overnight, resuspended in $^2\text{H}_2\text{O}$ and titrated to pD 3.0 with deuterated formic acid. Note that at pH values above 4.5, SIV e-gp41 precipitates out of solution. From ^1H - ^{15}N correlation spectra recorded at several different pH values, however, there is no evidence for any conformational change in SIV e-gp41 in the pH

range over which it remains soluble. NMR experiments were performed at 45°C on Bruker DMX500, DMX600 and DMX750 spectrometers equipped with x,y,z -gradient triple resonance probes. Spectral processing and analysis were performed using the programs NMRPipe (Delaglio *et al.*, 1995) and PIPP/CAPP (Garrett *et al.*, 1991), respectively.

NMR assignments, the secondary structure, and a topological model of SIV e-gp41. Given the high degree of sequence identity (~55%) between SIV and HIV e-gp41, we expect that the present results should be directly transferrable to HIV e-gp41.

To avoid problems associated with multiple potential modes of intra- and intersubunit disulfide bond formation, a double mutant of SIV e-gp41 was employed in which the two cysteine residues, Cys86 and Cys92, were substituted by alanine. We note, however, that these mutations had only a minimal effect on the structure of e-gp41 since the ^1H - ^{15}N correlation spectra of wild-type and mutant e-gp41 are virtually superimposable. The initial experiments were performed using a perdeuterated sample to reduce relaxation effects that severely limit NMR studies of large biomolecules (Yamazaki *et al.*, 1994; Grzesiek *et al.*, 1995; Venters *et al.*, 1995, 1996; Garrett *et al.*, 1997). In this regard we note that the average backbone amide T_2 relaxation time at 45°C is ~10 to 12 ms for the protonated sample but is increased to ~17–18 ms upon deuteration. The 2D ^1H - ^{15}N HSQC spectrum of perdeuterated SIV e-gp41 is shown in Figure 1A. Due to extensive spectral overlap, only 50% of the expected cross-peaks are well resolved in the 2D

spectrum. The majority of cross-peaks, however, can be resolved in the 3D heteronuclear experiments. Assignments were made using triple resonance experiments to delineate through-bond correlations along the protein backbone and side-chains (for reviews, see Clore & Gronenborn, 1991; Bax & Grzesiek, 1993; Gronenborn & Clore, 1995), and a summary of the experiments recorded is provided in Table 1. Using these procedures we were able to obtain complete ^1H , ^{15}N , ^{13}C and ^{13}CO backbone assignments which are provided as Supplementary Material. In addition, approximately 80% of the ^1H and ^{13}C side-chain assignments have been obtained. Particularly helpful in this regard was the use of a 4D ^{13}C - ^{13}C NOE experiment recorded at 750 MHz to delineate through-space ^{13}C - ^{13}C connectivities between directly bonded side-chain carbon atoms (Fisher *et al.*, 1996). An example of the quality of the data is illustrated in Figure 1B, which shows a series of strips from the 3D HNCA spectrum recorded on $^2\text{H}/^{15}\text{N}/^{13}\text{C}$ -labeled e-gp41, demonstrating sequential assignments for a portion of the N-terminal helix. The signal-to-noise ratio of this experiment is reasonably high, thereby rendering the assignment of e-gp41 feasible; the presence of additional corre-

Table 1. Summary of NMR experiments recorded on SIV e-gp41

Experiments	Correlation
A. Through-bond correlation experiments	
$^2\text{H}/^{15}\text{N}/^{13}\text{C}$ -labeled e-gp41	
d-HNCA	$^{13}\text{C}^\alpha(i-1)-^{15}\text{N}(i)-\text{H}(i)$
d-HNCO	$^{13}\text{CO}(i-1)-^{15}\text{N}(i)-\text{H}(i)$
d-HN(CO)CA	$^{13}\text{C}^\alpha(i-1)-^{15}\text{N}(i)-\text{H}(i)$
d-HN(CA)CO	$^{13}\text{CO}(i)-^{15}\text{N}(i)-\text{H}(i)$
d-HNCACB	$^{13}\text{C}^\alpha/\text{C}^\beta(i-1)-^{15}\text{N}(i)-\text{H}(i)$
d-HN(COCA)CB	$^{13}\text{C}^\beta(i-1)-^{15}\text{N}(i)-\text{H}(i)$
d-HN(CA)CB	$^{13}\text{C}^\beta(i-1)-^{15}\text{N}(i)-\text{H}(i)$
d-C(CC)(CO)NH	$^{13}\text{C}(i-1)-^{15}\text{N}(i)-\text{H}(i)$
$^{15}\text{N}/^{13}\text{C}$ -labeled e-gp41	
HBHA(CO)NH	$\text{H}^\beta/\text{H}^\alpha(i-1)-^{15}\text{N}(i)-\text{H}(i)$
HCACO	$^{13}\text{CO}(i)-^{13}\text{C}^\alpha(i)-\text{H}^\alpha(i)$
HNCO	$^{13}\text{CO}(i-1)-^{15}\text{N}(i)-\text{H}(i)$
^{15}N -labeled e-gp41	
HNHA	$\text{H}^\alpha(i)-^{15}\text{N}(i)-\text{H}(i)$
B. Through-space correlation experiments	
^{15}N -labeled e-gp41	
3D ^{15}N -separated NOE	$\text{H}-\text{HN}(^{15}\text{N})$
$^{15}\text{N}/^{13}\text{C}$ -labeled e-gp41	
3D ^{13}C -separated NOE	$\text{H}-\text{H}(^{13}\text{C})$
4D $^{13}\text{C}/^{15}\text{N}$ -separated NOE	$\text{H}(^{13}\text{C})-\text{H}(^{15}\text{N})$
4D HCCH ^{13}C - ^{13}C NOE	$\text{H}_j-^{13}\text{C}_j-^{13}\text{C}_{j\pm 1}-\text{H}_{j\pm 1}$
$^2\text{H}/^{15}\text{N}$ -labeled e-gp41	
4D $^{15}\text{N}/^{15}\text{N}$ -separated NOE	$\text{HN}(^{15}\text{N})-\text{HN}(^{15}\text{N})$
Mixed $^{15}\text{N}/^2\text{H}$ and ^{13}C -labeled heterotrimer	
3D ^{15}N -separated/ ^{13}C -filtered NOE	$\text{HN}(^{15}\text{N})-\text{H}(^{13}\text{C})$ intersubunit
3D ^{13}C -separated/ ^{15}N -filtered NOE	$\text{H}(^{13}\text{C})-\text{HN}(^{15}\text{N})$ intersubunit

d, denotes deuteration decoupling which was carried out as described by Garrett *et al.* (1997). All the experiments (or simple modifications thereof for ^2H -labeled samples) have either been described in the following three reviews (Clore & Gronenborn, 1991; Bax & Grzesiek, 1993; Gronenborn & Clore, 1995) or in the following papers: Yamazaki *et al.* (1994), HN(CA)CB and HN(COCA)CB; Wittekind & Mueller (1993), HNCACB; Grzesiek *et al.* (1995), 4D $^{15}\text{N}/^{15}\text{N}$ -separated NOE; Farmer & Venters (1995), C(CC)CONH; Clubb *et al.* (1992), HN(CA)CO; Fisher *et al.* (1996), 4D HCCH ^{13}C - ^{13}C NOE. The subscript i refers to the residue number, while the subscript j refers to the carbon position within the side-chain of any given residue.

lations, however, demonstrates some of the technical difficulties arising from extensive spectral overlap. Interestingly, a second set of correlations is observed for residues 85 to 104, revealing the presence of a minor form. The relative intensities of the cross-peaks arising from the major and minor forms indicate that the population of the minor form is ~10%. The presence of three proline residues in this region (Pro97, Pro99 and Pro105) suggests that the minor form may be due to proline *cis-trans* isomerization. This is supported by the observation that the ^{13}CO , $^{13}\text{C}^\alpha$ and $^{13}\text{C}^\beta$ chemical shifts of the major form of Pro99 (176.7, 63.2 and 31.1 ppm, respectively) are characteristic of the *trans* isomer, while those of the minor form of

Pro99 (175.7, 62.4 and 33.1 ppm, respectively) are characteristic of the *cis* isomer (Richarz & Wüthrich, 1978).

The secondary structure was delineated on the basis of short and medium-range NOEs, NH exchange rates, $^3J_{\text{HN}\alpha}$ coupling constants, and ^{13}C secondary shifts summarized in Figure 2. Two regions, corresponding to residues 29 to 77 and 108 to 147, exhibit a contiguous stretch of $\text{C}^\alpha\text{H}(i)\text{-NH}(i+3)$ NOEs, slowly exchanging backbone amides, and small (<7 Hz) $^3J_{\text{HN}\alpha}$ coupling constants, which are characteristic of helices (Wüthrich, 1986; Clore & Gronenborn, 1989). In addition, the ^{13}CO , $^{13}\text{C}^\alpha$ and $^{13}\text{C}^\beta$ resonances for these two regions exhibit secondary ^{13}C shifts

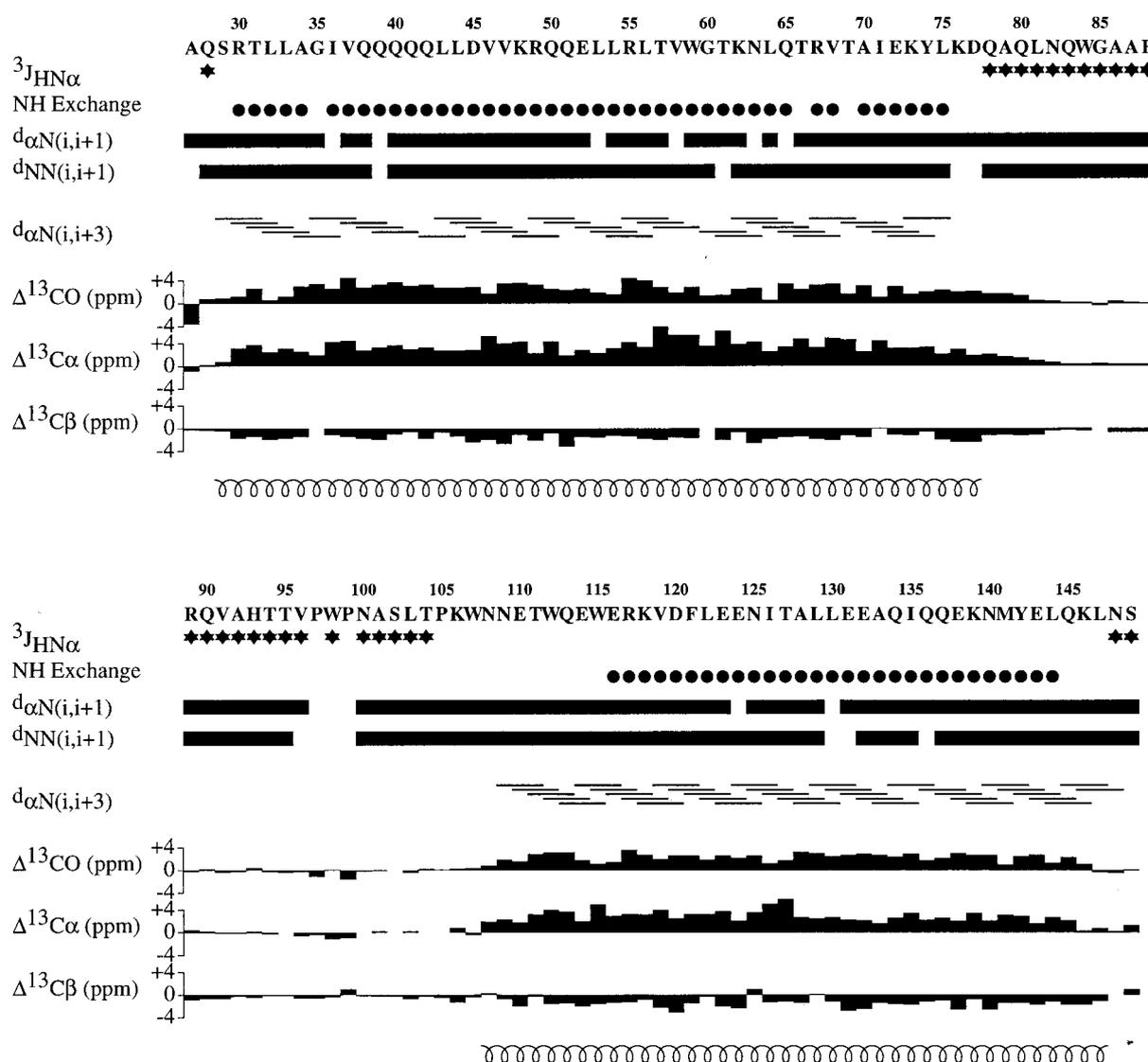


Figure 2. Summary of sequential and medium-range NOEs, $^3J_{\text{HN}\alpha}$ coupling constants, NH exchange rates, and secondary ^{13}CO , $^{13}\text{C}^\alpha$ and $^{13}\text{C}^\beta$ chemical shifts for SIV e-gp41. Stars represent $^3J_{\text{HN}\alpha}$ coupling constants ≥ 7 Hz. Filled-in circles represent NH exchange rates with a $t_{1/2} > 6$ hours. Secondary chemical shifts are relative to the random coil values of Wishart *et al.* (1995). NH exchange was estimated from a 3D HNC0 experiment which correlates the $\text{CO}(i-1)\text{-N}(i)\text{-HN}(i)$ resonances, thereby resolving the spectral overlap seen in the 2D $^1\text{H}\text{-}^{15}\text{N}$ correlation spectrum. The HNC0 experiment was initiated ~30 minutes after taking up a lyophilized sample of e-gp41 in $^2\text{H}_2\text{O}$ and recorded with a total acquisition time of 20 hours.

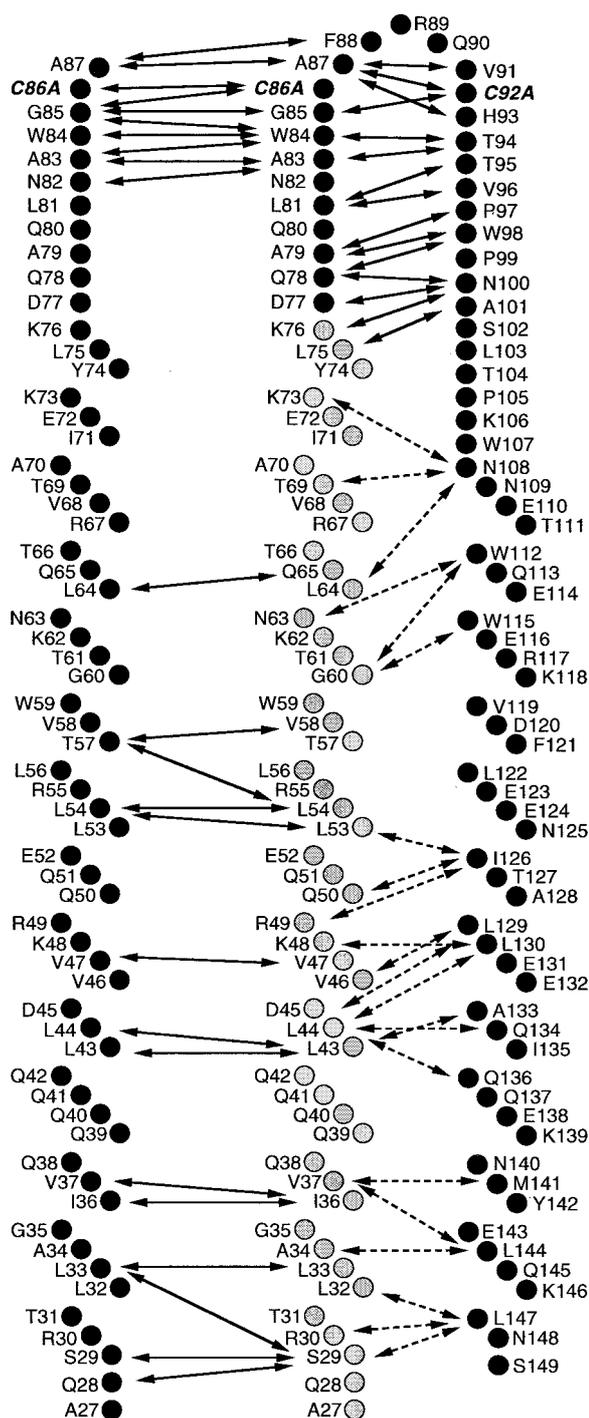


Figure 3. Summary of long-range NOEs observed for SIV e-gp41 in a 3D ^{15}N -separated NOE experiment (80 ms mixing time) recorded on a ^{15}N -labeled sample, 4D $^{13}\text{C}/^{15}\text{N}$ -separated NOE experiment (80 ms mixing time) recorded on a $^{15}\text{N}/^{13}\text{C}$ -labeled sample, and in 3D ^{13}C -separated/ ^{15}N -filtered and ^{15}N -separated/ ^{13}C -filtered NOE spectra (120 ms mixing time) recorded on a sample containing a mixture of $^{15}\text{N}/^2\text{H}/^{12}\text{C}$ and $^{13}\text{C}/^1\text{H}/^{14}\text{N}$ labeled subunits to identify inter-subunit NOEs. The location of the Cys86 \rightarrow Ala and Cys92 \rightarrow Ala mutations are shown in bold and italics. In the case of the NOEs indicated by broken lines between the N and C-terminal helices, no distinction between intra and inter-subunit NOEs has been made at the present level of analysis.

that are typical of a helix (i.e. the ^{13}CO and $^{13}\text{C}^\alpha$ resonances are downfield shifted while the $^{13}\text{C}^\beta$ resonances are upfield shifted, relative to their random coil expectation values; Spera & Bax, 1991; Wishart *et al.*, 1995). In contrast, residues 78 to 107 and the N (residues 27 to 28) and C (residues 148 to 149) termini exhibit fast NH exchange rates, relatively large $^3J_{\text{HN}\alpha}$ coupling constants, and ^{13}C secondary shifts that are near random coil values, characteristic of regions that do not possess regular secondary structure.

Taken together, the present results indicate that SIV e-gp41 consists of two long helices, a 49-residue helix at the N terminus (residues 29 to 77) and a 40-residue helix at the C terminus (residues 108 to 147), connected by a 30-residue loop (residues 78 to 107). We also note that residues 79 to 105 exhibit cross-peak intensities in the 3D HNCO spectrum that are 5 to 25 times more intense than those of the helical regions (data not shown), suggesting that the loop connecting the two helices is more mobile than the helical core.

Carrying out a full three-dimensional structure determination of SIV e-gp41 by NMR presents a significant challenge due to its large molecular weight, limited spectral dispersion (as a result of the high helical content and complete absence of any β -sheet) and trimeric nature. Even at the present stage of analysis, however, it is possible to derive information concerning the tertiary and quaternary structures. Figure 3 displays a number of long range intra and inter-subunit NOEs that have been assigned at the present time and can be used to deduce an approximate global fold of e-gp41. The current set of intersubunit NOEs were identified by recording 3D ^{15}N -separated/ ^{13}C -filtered and ^{13}C -separated/ ^{15}N -filtered NOE spectra on a sample comprising a mixture of $^{15}\text{N}/^2\text{H}/^{12}\text{C}$ and $^{13}\text{C}/^1\text{H}/^{14}\text{N}$ labeled monomers. A qualitative analysis of the NOE data clearly indicates that the N and C-terminal helices are oriented antiparallel with respect to each other, consistent with a previous model derived from protease protection (Blacklow *et al.*, 1995; Lu *et al.*, 1995) and EPR studies (Rabenstein & Shin, 1996). Moreover, the N-terminal region of the C-terminal helix exhibits intramolecular NOEs with the C-terminal region of the N-terminal helix, whereas the C-terminal region of the C-terminal helix displays intermolecular NOEs with the N-terminal region of the N-terminal helix (Figure 3). This suggests that the C-terminal helix is oriented somewhat obliquely to the N-terminal helices. The N-terminal helices, on the other hand, are oriented approximately parallel to each other. No intersubunit NOEs have yet been identified between the C-terminal helices. The loop connecting the N and C-terminal helices displays numerous long range NOEs, both intra and intermolecular, indicating that the loops from the three subunits interact with each other and form a well-defined structure (Figure 3). In addition, Ala86 and Ala92, which correspond to the conserved cysteine residues (Cys86 and Cys92) in the wild-type sequence, are in close spatial proximity between subunits, consistent with the presence of intersubunit disulfide bonds observed in SIV (Wingfield *et al.*, 1997) and HIV (Weissenhorn *et al.*, 1996) e-gp41.

Equilibrium centrifugation studies have demonstrated that SIV e-gp41 exists as a trimer at pH values below 4 (Wingfield *et al.*, 1997). The presence of a single set of correlations in the NMR spectra (with the exception of the minor form present in the loop) indicates that the complex is symmetric. Based on the present observations, a model for e-gp41 is presented in Figure 4. The N-terminal helices of the three subunits are packed within the protein interior in a parallel trimeric coiled-coil arrangement. The C-terminal helices are located on the outside of the trimer with each C-terminal helix oriented antiparallel to two adjacent N-terminal helices. Hence, the loops connecting the N and C-terminal helices are located on the same side of the molecule.

This model is compatible with a number of previous studies. First, intersubunit cross-linking of

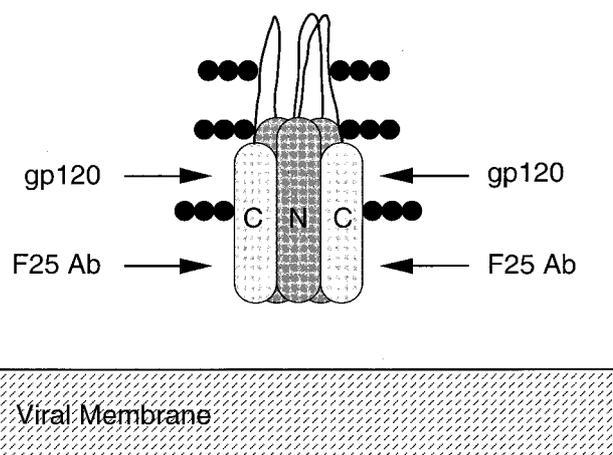


Figure 4. Schematic diagram of the NMR-derived topological model of trimeric SIV e-gp41. Filled-in circles represent glycosylation sites, arrows the binding sites for gp120 and the monoclonal antibody 2F5. The three N-terminal helices are located within the protein interior and form a parallel trimeric coiled-coil. The three C-terminal helices are arranged antiparallel to the N-terminal helices and are located on the outside of the molecule. Only two of the three C-terminal helices are visible in the schematic diagram; the third C-terminal helix is located at the back of the molecule and hence is not visible in this view.

the conserved cysteine residues has been observed in HIV and SIV e-gp41 (Weissenhorn *et al.*, 1996; Wingfield *et al.*, 1997), consistent with the proximal location of the loop regions in the model. Second, the putative glycosylation sites at Asn100, Asn109 and Asn125 (Dedra *et al.*, 1992) are located in the loop region and the C-terminal helix, which are exposed to solvent in the model. Third, mutagenesis studies of HIV gp41 have suggested that gp120 binds near Val119 (Kowalski *et al.*, 1989), which is located in the exposed C-terminal helix. Fourth, the binding site for the neutralizing monoclonal antibody 2F5 is centered around Tyr142 (Muster *et al.*, 1993), which is located in the exposed C-terminal helix. Interestingly, the observation that 2F5 binds to the gp41/gp120 complex suggests that gp120 does not interact with the region in the vicinity of Tyr142 of e-gp41. Finally, the overall dimensions of the model complex are similar to that observed in electron micrographs of HIV e-gp41 (Weissenhorn *et al.*, 1996). Thus, the overall length of SIV e-gp41 can be estimated as the length of the N-terminal helix plus residues 78 to 89 of the loop region (cf. Figures 3 and 4). Considering that the interresidue distance is 1.5 Å in a helix and ~4 Å in a loop region, the overall length is calculated to be ~122 Å, in good agreement with the value of 125(±13) Å determined by electron microscopy (Weissenhorn *et al.*, 1996).

Just prior to submission of this Communication, a 2.0 Å resolution X-ray structure of a trimer of two interacting peptides, N36 and C34, derived from

HIV gp41 and corresponding to residues 34 to 69 and 112 to 145, respectively, of SIV gp41, was published (Chan *et al.*, 1997). The crystal structure therefore comprises only ~60% of the SIV e-gp41 construct used in the present work. Both peptides are entirely helical in the X-ray structure, consistent with the NMR data, and the overall topology of the trimer observed crystallographically is essentially the same as that of the model presented in Figure 4. The present NMR data, however, indicate that both the N and C-terminal helices extend significantly beyond the peptides used in the X-ray work. Thus, the N-terminal helix (residues 29 to 77) is in fact ~36% longer than the N36 peptide, while the C-terminal helix (residues 108 to 147) is ~18% longer than the C34 peptide. Moreover, our data indicate that the loop connecting the two helices is not disordered but adopts a well-defined structure in which the loops from the three subunits interact with each other (Figure 3). We are currently in the process of determining a complete three-dimensional structure of the SIV e-gp41 trimer, principally exploiting multidimensional isotope-edited and filtered NOE experiments. In this manner, we will elucidate the determinants of protein-protein interaction as exemplified by the helix and loop packing of the e-gp41 trimer. In addition, a complete structural description of e-gp41 may provide further insight into the role of gp41 in HIV and SIV infection.

Acknowledgements

We thank D. Garrett and F. Delaglio for software support, R. Tschudin for hardware support, and A. Bax, N. Tjandra and J. Shu 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 and A.M.G).

References

- Allan, J. S., Coligan, J. E., Barin, F., McLane, M. F., Sodroski, J. G., Rosen, C. A., Haseltine, W. A., Lee, W. A. & Essex, M. (1985). Major glycoprotein antigens that induce antibodies in AIDS patients are encoded by HTLV-III. *Science*, **228**, 1091–1094.
- Bax, A. & Grzesiek, S. (1993). Methodological advances in protein NMR. *Acc. Chem. Res.* **26**, 131–138.
- Blacklow, S. C., Lu, M. & Kim, P. S. (1995). A trimeric subdomain of the simian immunodeficiency virus envelope glycoprotein. *Biochemistry*, **34**, 14955–14962.
- Chan, D. C., Fass, D., Berger, J. M. & Kim, P. S. (1997). Core structure of gp41 from the HIV envelope glycoprotein. *Cell*, **89**, 263–273.
- Clore, G. M. & Gronenborn, A. M. (1989). Determination of three-dimensional structures of proteins and nucleic acids in solution by nuclear magnetic resonance spectroscopy. *CRC Crit. Rev. Biochem. Mol. Biol.* **24**, 479–564.
- Clore, G. M. & Gronenborn, A. M. (1991). Structures of larger proteins in solution: three- and four-dimensional heteronuclear magnetic resonance spectroscopy. *Science*, **252**, 1390–1399.
- Clubb, R. T., Thanabal, V. & Wagner, G. (1992). A constant-time three-dimensional triple-resonance pulse scheme to correlate intraresidue $^1\text{H-N}$, ^{15}N , and ^{13}C chemical shifts in ^{15}N - ^{13}C labeled proteins. *J. Magn. Reson.* **97**, 213–217.
- Dagleish, A. G., Beverley, P. C. L., Clapham, P. R., Crawford, D. H., Greaves, M. F. & Weiss, R. A. (1984). The CDA(T4) antigen is an essential component of the receptor for the AIDS retrovirus. *Nature*, **312**, 763–767.
- Dedra, D., Gu, R. & 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. & Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR*, **6**, 277–293.
- Deng, H., Liu, R., Ellmeier, W., Choe, S., Unutmaz, D., Burkhart, M., Di Marzio, P., Marmon, S., Sutton, R. E., Hill, C. M., Davis, C. B., Peiper, S. C., Schall, T. J., Littman, D. R. & Landau, N. R. (1996). Identification of a major co-receptor for primary isolates of HIV-1. *Nature*, **381**, 661–666.
- Farmer, B. T., II & Venters, R. A. (1995). Assignment of side-chain ^{13}C resonances in perdeuterated proteins. *J. Am. Chem. Soc.* **117**, 4187–4188.
- Feng, Y., Border, C. C., Kennedy, P. E. & Berger, E. A. (1996). HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G-protein-coupled receptor. *Science*, **272**, 872–877.
- Fisher, M. W. F., Zeng, L. & Zuiderweg, E. R. P. (1996). Use of ^{13}C - ^{13}C NOE for the assignment of NMR lines of larger labeled proteins at larger magnetic fields. *J. Am. Chem. Soc.* **118**, 12457–12458.
- Gallaher, W. R. (1987). Detection of a fusion peptide sequence in the transmembrane proteins of human immunodeficiency virus. *Cell*, **50**, 327–328.
- Garrett, D. S., Powers, R., Gronenborn, A. & 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.* **95**, 214–220.
- Garrett, D. S., Seok, Y.-J., Liao, D.-I., Peterkofsky, A., Gronenborn, A. M. & Clore, G. M. (1997). Solution structure of the 30 kDa N-terminal domain of enzyme I of the *Escherichia coli* phosphoenolpyruvate:sugar phosphotransferase system by multidimensional NMR. *Biochemistry*, **36**, 2517–2530.
- Gronenborn, A. M. & Clore, G. M. (1995). Structures of protein complexes by multidimensional heteronuclear magnetic resonance spectroscopy. *CRC Crit. Rev. Biochem. Mol. Biol.* **30**, 351–385.
- Grzesiek, S., Wingfield, P. T., Stahl, S. J., Kaufman, J. D. & Bax, A. (1995). Four-dimensional ^{15}N -separated NOESY of slowly tumbling perdeuterated ^{15}N -enriched proteins. *J. Am. Chem. Soc.* **117**, 9594–9595.
- Klatzmann, D. E., Champagne, E., Chamaret, S., Gruet, J., Guetard, D., Herceud, J., Gluckman, J. C. & Montagnier, L. (1984). T-lymphocyte T4 molecule behaves as the receptor for human retrovirus LAV. *Nature*, **312**, 767–768.
- Kowalski, M., Potz, J., Basirpour, L., Dorfman, T., Goh, W. C., Terwilliger, E., Dayton, A., Rosen, C., Haseltine, W. & Sodroski, J. (1987). Functional regions of the envelope glycoprotein of human

- immunodeficiency virus type I. *Science*, **237**, 1351–1355.
- Lawless, M. K., Barnery, S., Gutherie, K. I., Bucy, T. B., Petteway, S. R. & Merutka, G. (1996). HIV-1 membrane fusion mechanism: structural studies of the interactions between biologically-active peptides from gp41. *Biochemistry*, **35**, 13697–13708.
- Litwin, D. T., Allaway, G. P., Martin, S. R., Huang, Y., Nagashima, K. A., Cayanan, C., Maddon, P. J., Koup, R. A., Moore, J. P. & Paxton, W. A. (1996). HIV-1 entry into CD4⁺ cells is mediated by the chemokine receptor CC-CKR-5. *Nature*, **381**, 667–673.
- Lu, M., Blacklow, S. C. & Kim, P. S. (1995). A trimeric structural domain of the HIV-1 transmembrane glycoprotein. *Nature Struct. Biol.* **2**, 1075–1082.
- Muster, T., Steindl, F., Purtscher, M., Trkola, A., Klima, A., Himmler, G., Rucker, F. & Katinger, H. (1993). A conserved neutralizing epitope on gp41 of human immunodeficiency virus type I. *J. Virol.* **67**, 6642–6647.
- Rabenstein, M. D. & Shin, Y.-K. (1996). HIV-1 gp41 tertiary structure studied by EPR spectroscopy. *Biochemistry*, **35**, 13922–13928.
- Richarz, R. & Wüthrich, K. (1978). Carbon-13 NMR chemical shifts of the common amino acid residues measured in aqueous solutions of the linear tetrapeptides H-Gly-Gly-X-L-Ala-OH. *Biopolymers*, **17**, 2133–2141.
- Spera, S. & Bax, A. (1991). An empirical correlation between protein backbone conformation and C α and C β chemical shifts. *J. Am. Chem. Soc.* **113**, 5490–5492.
- Venters, R. A., Metzler, W. J., Spicer, L. D., Mueller, L. & Farmer, B. T., II (1995). Use of ¹H_N–¹H_N NOEs to determine protein global folds in perdeuterated proteins. *J. Am. Chem. Soc.* **117**, 9592–9593.
- Venters, R. A., Farmer, B. T., II, Fierke, C. A. & Spicer, L. D. (1996). Characterizing the use of perdeuteration in NMR studies of larger proteins: ¹³C, ¹⁵N and ¹H assignments of human carbonic anhydrase II. *J. Mol. Biol.* **264**, 1101–1116.
- Veronese, F. D., DeVico, A. L. D., Copeland, T. D., Oroszlan, S., Gallo, R. C. & Sarngadharan, M. G. (1985). Characterization of gp41 as the transmembrane protein coded by HTLV-III/LAV envelope gene. *Science*, **229**, 1402–1405.
- Weissenhorn, W., Wharton, S. A., Calder, L. J., Earl, P. L., Moss, B., Aliprandis, E., Skehel, J. J. & Wiley, D. C. (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.
- Wingfield, P. T., Stahl, S. J., Kaufman, J., Zlotnick, A., Hyde, C. C., Gronenborn, A. M. & Clore, G. M. (1997). The extracellular domain of immunodeficiency virus gp41 protein: expression in *Escherichia coli*, purification and crystallization. *Protein Sci.*, in the press.
- Wishart, D. S., Bigam, C. G., Hol, A., Hodges, R. S. & Sykes, B. D. (1995). ¹H, ¹³C and ¹⁵N random coil NMR chemical shifts of the common amino acids. *J. Biomol. NMR*, **5**, 67–81.
- Wittekind, M. & Mueller, L. (1993). HNCACB, a high-sensitivity 3D NMR experiment to correlate amide protons and nitrogen resonances with the alpha- and beta-carbon resonances in proteins. *J. Magn. Reson.* **101**, 201–205.
- Wüthrich, K. (1986). *NMR of Proteins and Nucleic Acids*, Wiley, New York.
- Yamazaki, T., Lee, W., Arrowsmith, C., Muhandiram, D. & Kay, L. E. (1994). A suite of triple resonance NMR experiments for the backbone assignment of ¹⁵N, ¹³C, ²H labeled proteins with high sensitivity. *J. Am. Chem. Soc.* **116**, 11655–11666.

Edited by P. E. Wright

(Received 25 April 1997; received in revised form 13 June 1997; accepted 17 June 1997)



<http://www.hbuk.co.uk/jmb>

Supplementary material comprising a Table of ¹H, ¹³C, ¹³CO and ¹⁵N resonance assignments of SIV e-gp41 is available on JMB Online.