The HIV envelope glycoprotein gp120 plays a critical role in virus entry and thus its structure is of extreme interest for the development of novel therapeutics and vaccines. To date, high resolution structural information about gp120 in complex with gp41 has proven intractable. In this study we characterize the structural properties of gp120 in the presence and absence of gp41 domains by NMR. Using the peptide probe 12p1 (sequence = RINNPWSEAMM), which was previously identified as an entry inhibitor that binds to gp120, we identify atoms of 12p1 in close contact with gp120 in the monomeric and trimeric states. Interestingly the binding mode of 12p1 with gp120 is similar for clades B and C. In addition, we show a subtle difference in the binding mode of 12p1 in the presence of gp41 domains, i.e. the trimeric state, which we interpret as small differences in the gp120 structure in the presence of gp41. The envelope glycoproteins gp120 and gp41 form a non-covalent complex on the surface of HIV and mediate entry (1). Gp120 initiates viral entry by binding to receptors found on the target cell surface, CD4 and a chemokine co-receptor, and gp41 subsequently brings the target and viral membranes in close proximity, thereby allowing membrane fusion and entry. Importantly, conformational changes of gp120 and gp41 are thought to be intrinsic to the entry process (2-4). Due to their critical role in HIV entry, gp120 and gp41 have been intensively studied by x-ray crystallography and NMR spectroscopy (5). For example there is structural information for the gp120 core in the free state and in complex with CD4 receptor domains and various monoclonal antibodies (6-8) and there are numerous structures of the gp41 extracellular domain in the fusion state (9-12). However, all gp120 structures to date are missing significant regions of gp120, and perhaps more importantly, domains of gp41 (5).
a consequence, there exists the possibility that the gp120 structure in complex with gp41, the state present on virus and HIV infected cells, is structurally different than those previously characterized.

The critical role of gp120 has led to efforts to find molecules that bind to gp120 and inhibit its interaction with receptors (5). For example, the peptide 12p1 (sequence = RINNIPWSEAMM), which was discovered by phage display against recombinant HIV-1 HXB2 gp120, binds to gp120 and blocks its interaction with CD4 with an IC50 ranging from 0.3 to 25 µM for different strains (13). Subsequent studies established that 12p1 binds to gp120 with a stoichiometry of 1:1 and inhibits viral entry with an IC50 of ~8 µM (14). As a consequence, 12p1 and its derivatives present the potential for development as HIV entry inhibitors (15,16). In the present work we have exploited 12p1 to probe gp120 structure in the presence and absence of gp41 domains by NMR. Specifically, we have assigned the 12p1 1H, 13C and 15N resonances and employed Saturation Transfer Difference (STD) NMR (17) to identify 12p1 1H in close contact with the gp120 binding surface.

**Experimental Procedures**

*Preparation of 12p1 and recombinant HIV envelope glycoproteins*

12p1 was synthesized by solid phase peptide synthesis with unmodified N- and C-terminal groups. The peptide was purified by HPLC and the mass was verified by MALDI-TOF mass spectrometry. In our initial work we found that the tryptophan indole was especially sensitive to oxidation, as previously observed for other peptides and proteins (18), and thus the peptide solid was stored under vacuum and all NMR solutions were stored in the presence of argon to avoid oxidation. For preparation of envelope constructs from strain R2, a codon optimized R2 Env gene was cloned into the vector phCMVhygro, which was modified from the plasmid phCMV (Genlantis, Inc) by introduction of a hygromycin resistance gene (phCMV is a CMV promoter and enhancer driven cassette). The R2 Env gene was truncated at the transmembrane domain and the arginine residues at both gp120-gp41 cleavage sites where mutated to serine (R517S, R520S) and the GCN4 trimerization domain was fused to the carboxy terminus of R2gp140 to form R2gp140-GCN, as described by Yang et al. (19,20). Monomeric R2gp120 (molecular weight 120 kDa by SDS-PAGE) was produced as previously described (21), the R2gp140 envelopes were produced by stable cell line expression. All R2 envelopes were prepared by expression in serum-free conditions using 293T cells, and purified using a sequential combination of DEAE adsorption, lentil lectin affinity, and size exclusion chromatography. The purity of the recombinant R2 envelope preparations was assessed by SDS-PAGE. The oligomerization state of the R2gp140 envelopes were assessed by Blue Native PAGE; R2gp140-GCN and R2gp140-linker-GCN migrate as
trimers (molecular weight 720 kDa). The conformational integrity of the recombinant R2 envelopes were demonstrated by their binding to soluble CD4, CCR5, and a series of monoclonal antibodies. Gp120 from strain YU2 was obtained from Immune Technology Corporation. Gp120 from strain 96ZM65 was obtained from the NIH AIDS Research and Reference Reagent Program. Note that all gp120 and gp140 constructs produced in eukaryotic cells are expected to be fully glycosylated.

NMR Experiments
NMR experiments were performed on Bruker 600, 800 and 900 MHz spectrometers equipped with room temperature or cryogenic temperature triple resonance probes. Buffer conditions were either 50 mM PO₄/pH 7.2 or PBS (20 mM PO₄/pH 7.2, 150 mM NaCl) in 90% H₂O, 10% ²H₂O or 100% ²H₂O at 25°C. NMR assignments were obtained by a series of TOCSY, ROESY (mixing time=100 msec), NOESY (mixing time=200 msec), ¹³C-edited HSQC, and ¹⁵N-edited HSQC (22,23). STD NMR experiments were performed with a 50 msec gaussian-shaped saturating pulse for 2.5 sec with “on” resonance saturation at -1.5 ppm and “off” resonance saturation at 30 ppm. All data were processed by NMRPipe (24) with referencing to the water peak at 4.773 ppm and using a 10 Hz line-broadening window.

RESULTS
Assignment of 12p1 resonances
As a first step, we performed a series of NMR experiments to assign the ¹H, ¹³C and ¹⁵N resonances of 12p1 as a free peptide. Examples of the TOCSY amide region and ¹³C-edited HSQC spectra with assignments are shown in Figure 1a, 1b and 1c (examples of the TOCSY aromatic region and ¹⁵N-edited HSQC are shown in Figure S1 of the Supplementary Information). We note that the ¹³C-edited HSQC was especially useful in the assignment process due to the significant overlap of 12p1 ¹H. In total, assignments were obtained for over 95% of the 12p1 ¹H, ¹³C and ¹⁵N (see Table S1 of the Supplementary Information). It was next of interest to assess the secondary structure of the peptide from secondary chemical shift differences, which are defined as δ_{observed}-δ_{random coil} (25). As shown in Figure 1d, the secondary chemical shifts of the ¹Hα and ¹³Cα indicate that 12p1 exhibits the absence of regular secondary structure, which is not surprising for a peptide. Interestingly, P6 exhibits chemical shifts indicative of a trans conformation. For example, the observed ¹Hα, ¹³Cα, ¹Hβ, and ¹³Cβ of P6 are 4.36, 63.4, 2.21/1.80, and 31.9 ppm, respectively (Table S1 of Supplementary Information), which are closer to the values observed for trans proline (the random coil values of the ¹Hα, ¹³Cα, ¹Hβ, and ¹³Cβ of trans proline are 4.45, 63.7, 2.29/1.99, and 32.2 ppm, respectively, and the random coil values of the ¹Hα, ¹³Cα, ¹Hβ, and ¹³Cβ of cis proline are 4.60, 63.0, 2.39/2.18, and 34.8 ppm, respectively, 25).
NMR study of the 12p1 interaction with monomeric gp120

STD NMR presents a powerful method to characterize interactions between small molecules and large molecules (17). In the STD NMR experiment, the resonances of the large molecule $^1$H are selectively irradiated, subsequently magnetization is transferred to the $^1$H of small molecules that exchange between bound and free states during the irradiation period; the difference, with respect to a reference spectrum in which the large molecule $^1$H are not irradiated and hence no magnetization transfer occurs, identifies $^1$H in closest contact in the bound state. The 1D $^1$H NMR reference spectrum of the 12p1 downfield region is shown in Figure 2a with the W7 sidechain $^1$H identified. In Figure 2b, the STD NMR spectrum of 12p1 in the absence of gp120 exhibits the absence of signal as expected. In contrast, in Figure 2c, the STD NMR spectrum of 12p1 in the presence of gp120 from strain YU2, which is a monomer of 120 kDa, suggests that in the bound state the W7 sidechain atoms $H_{\delta_1}$, $H_{\zeta_2}$, $H_{\eta_2}$, $H_{\xi_3}$ and $H_{\zeta_3}$ of 12p1 are in close contact with gp120. Interestingly, the absence of other resonances in this spectral region (i.e., the backbone and sidechain $^1$H$_N$ present in Figure 2a) indicates that these $^1$H are not in close contact to gp120. In Figures 2d and 2e, similar STD NMR spectra are observed for 12p1 in the presence of gp120 from strains R2 and 96ZM65, respectively, suggesting similar interactions between 12p1 and gp120 in a strain independent manner. Finally, we note that STD NMR signals are also observed in the aliphatic region of 12p1, which is also indicative of binding; however, due to spectral overlap, specific interactions are difficult to identify (Figure S2 of the Supplementary Information).

NMR study of the 12p1 interaction with trimeric gp140

It is next of interest, to contrast the 12p1 interaction with gp120 in the presence and absence of gp41 domains. For this part of the study, 2 more constructs were tested, as depicted in Figure 3a. The first construct, denoted R2gp140-GCN, consisted of full length R2 strain gp120 followed by gp41 in which the C-terminal transmembrane and cytoplasmic domains have been deleted. In this construct, the furin cleavage site has been mutated by arginine to serine substitutions and thus gp120 is covalently attached to gp41 in order to stabilize the gp120-gp41 interaction (19,20). In addition, a GCN4 trimerization domain has been appended to the C-terminus of gp41 to stabilize the trimeric state (19,20). The second construct, denoted R2gp140-linker-GCN, is similar to R2gp140-GCN except that a 15 residue flexible linker domain has been inserted between the gp120-gp41 interface with the rationale that the linker region would remove possible conformational constraint present in the gp140 construct. In both cases, the purified R2gp140 constructs were trimers of 720 kDa. As shown in Figure 3b, the STD NMR spectrum of 12p1 in the presence of gp140 suggests that the
sidechain atoms of W7 are in close contact to trimeric gp120; however, the additional signal observed for the W7 indole group (H$_{\varepsilon}$) clearly indicates an additional interaction site with respect to monomeric gp120. This notion is further supported by the STD NMR spectrum of 12p1 in the presence of gp140-linker, as shown in Figure 3c. Interestingly, the similar spectra of the gp140 and gp140-linker also suggest that the presence of the linker does not measurably affect the interaction between 12p1 and trimeric gp120.

**DISCUSSION**

In this study, we first assigned the $^1$H, $^{13}$C and $^{15}$N resonances of 12p1, a 12 residue peptide that has shown promise as an inhibitor of HIV entry (13-16). Based on the chemical shifts of the free peptide, we find little evidence of secondary structure in the unbound state; however, P6 is predominantly in a trans conformation. Interestingly, the central seven residues of 12p1 comprised of N3-S8, which include P6, have been shown by single site alanine substitutions to be critical for the interaction with gp120 (13). As a consequence, the conformational state of P6 is thus expected to be of importance to binding, although at this point we do not have any evidence for preferential binding of gp120 to the cis or trans conformations of the 12p1 proline.

In a next step we demonstrated by STD NMR that the W7 sidechain of 12p1 was in close contact to gp120 from strains YU2, R2 and 96ZM65. Nonetheless, it is important to note that the studies presented herein were performed at a probe concentration of 1 mM to optimize signal/noise. An obvious limitation of the STD technique is the possibility of lower affinity interactions present at relatively high probe concentrations. Importantly, additional experiments by our group at 200 µM 12p1 exhibited a similar pattern in the STD spectrum, albeit at lower signal/noise, supporting the role of the 12p1 W7 sidechain in the peptide-gp120 interaction. The significance of the W7 sidechain to the 12p1-gp120 interaction is further supported by previous studies of W7A and W7F versions of 12p1 that exhibited greatly reduced binding to gp120, as measured by competition ELISA (13,14). Interestingly, the STD NMR spectra are similar for gp120 from clade B strains YU2 and R2 and clade C strain 96ZM65, which implies that the 12p1 interaction with gp120 occurs across multiple strains and clades of HIV-1. Previous studies have suggested that 12p1 binds to and inhibits viral entry of various clade B strains (13,14); however, the potential application to clade C, which is the most rapidly spreading subtype and is prevalent in Southern and East Africa, as well as in India (26), makes 12p1 even more attractive for development as a therapeutic.

To probe for structural differences between gp120 in the presence and absence of gp41 domains, we characterized the interaction between 12p1 and gp140, which is a surrogate for the gp120/gp41 trimeric state (19,20).
As noted in the introduction, there is no structural information for gp120 in the trimeric state (5), which is the physiological state present on virus and HIV infected cells. Interestingly, the STD NMR studies suggest that the W7 sidechain of 12p1 interacts with gp140 in a slightly different manner, specifically, the imino group H_{ε1} appears to be in closer contact to the gp120 binding surface. Moreover, the observation that a similar interaction occurs in the gp140 construct containing a linker, designed to remove potential constraint imposed by the absence of furin cleavage, suggests that the linker does not significantly perturb the 12p1 interaction site, an observation that could be of interest in the design of an HIV-1 envelope immunogen. In Figure 4a, a model for the differences between the 12p1 interaction with gp120 in the presence and absence of gp41 domains is presented to highlight the closer contact of the 12p1 H_{ε1} to gp120 in the presence of gp41 domains due to alteration of the binding pocket. These subtle differences in the interaction of 12p1 suggest subtle differences in the structure of gp120 in the presence of gp41 domains. The sensitivity of the interaction is further supported by previous work demonstrating that the addition of non-natural hydrophobic moieties to P6 significantly enhances the 12p1 affinity for gp120 (15,16).

In Figure 4b, we present a model for the interaction between 12p1 and gp120. A previous study of gp120 escape mutants suggested that the 12p1 interaction occurs near residues K97, E102 and R476 of gp120 (14), residues that are highly conserved across HIV-1 strains (27). As noted previously (14), the CD4 and CR binding sites are proximal but not overlapping with this region as is evident in Figure 4b. Based on mutation studies (5,28-30), the interaction site with gp41 is also proximal but not overlapping with the putative 12p1 binding site, an observation that is in agreement with the subtle differences observed in our STD NMR studies presented herein. As noted above, hydrophobicity at residues 6 and 7 of 12p1 appear to be important for its interaction with gp120. Accordingly, we have highlighted the exposed hydrophobic regions on the gp120 surface in Figure 4b. Interestingly, W96 of gp120, which is highly conserved across HIV-1 strains (27), is relatively exposed and in close proximity to the sites of the escape mutants; thus, W96 may present a potential interaction partner with the hydrophobic groups of 12p1 and its derivatives. Furthermore, R1 and E9 of 12p1 clearly present potential for stabilizing electrostatic interactions with K97, E102 and R476 of gp120.

In summary, we have shown that NMR presents the potential to map gp120 binding sites with a high degree of information. We are presently applying the STD NMR technique for studying the interactions of other small molecules with gp120 and gp41 with the goal of probing HIV envelope structure in different conformational states. In addition, we are currently extending our studies to other envelope proteins including those of Ebola, Influenza and SARS Coronavirus.
Importantly, the technique of STD NMR may in principle be applied to study of even larger systems such as whole virus or cells (31,32).
REFERENCES

FOOTNOTES

This work was partially supported by NIH grants AI066709, AI077767, and the Chicago D-CFAR (AI082151). In addition, we acknowledge support of the UIC Center for Structural Biology.

FIGURE LEGENDS

Fig. 1 Assignment and secondary structure of free 12p1. (a) Amide region of the TOCSY. (b) Aliphatic region of the $^{13}$C-edited HSQC at natural abundance. (c) Aromatic region of the $^{13}$C-edited HSQC at natural abundance. (d) Secondary chemical shifts of 12p1. Random coil values were taken from Schwarzinger et al. (25). The experimental conditions for the TOCSY were 2 mM 12p1 in 20 mM PO$_4$/pH 7.2, 90% $^1$H$_2$O, 10% $^2$H$_2$O; the experimental conditions for the $^{13}$C-edited HSQC were 2 mM 12p1 in 50 mM PO$_4$/pH 7.2, 100% $^2$H$_2$O.

Fig. 2 STD NMR spectra of 12p1 in the presence of HIV gp120 constructs. (a) Reference 1D NMR spectrum of 12p1. (b) STD NMR spectrum of 12p1 in the absence of gp120. (c) STD NMR spectrum of 12p1 in the presence of gp120 from strain YU2. (d) STD NMR spectrum of 12p1 in the presence of gp120 from strain R2. (e) STD NMR spectrum of 12p1 in the presence of gp120 from strain 96ZM65. The experimental conditions were 1 mM 12p1 +/- 4 µM gp120 in 20 mM PO$_4$/pH 7.2, 90% $^1$H$_2$O, 10% $^2$H$_2$O.
Fig. 3  STD NMR of 12p1 in the presence of HIV gp140 constructs.  (a) Schematic diagram of the gp140 constructs used in the present study.  (b) STD NMR spectra of 12p1 in the presence of R2gp140-GCN.  (c) STD NMR spectrum of 12p1 in the presence of R2gp140-linker-GCN.  Experimental conditions were 1 mM 12p1 +/- 4 μM gp140 in 20 mM PO₄/pH 7.2, 90% ¹H₂O, 10% ²H₂O.

Fig. 4  Models for the 12p1 binding to gp120.  (a) Schematic model for the interaction between the W7 sidechain of 12p1 and the gp120 binding surface in the monomeric and trimeric states.  (b) Space filling representations for the interaction between 12p1 and gp120.  The structure of 12p1 represents a random coil model with P6 in the trans conformation, as determined by CNS (c.f. 12) and the gp120 structure is taken from Pancera et al. (8).  The exposed hydrophobic sidechains are colored green and select acidic and basic sidechains are colored, red and blue, respectively.  For reference, the bindings sites of CD4, CR and gp41 binding are colored purple, pink and yellow, respectively (5).
Figure 2
Figure 3

(a) gp120 and gp41 domains with Furin Site indicated.

R2 gp120

R2 gp140-GCN

R2 gp140-linker-GCN

(b) NMR spectra for R2 gp120.

(c) NMR spectra for R2 gp140-GCN.
Figure 4