Title

Structure and Interactions of HIV-1 gp41 CHR-NHR Reverse Hairpin Constructs Reveal Molecular Determinants of Antiviral Activity

Journal

Journal of Molecular Biology, 436(16)

ISSN

0022-2836

Authors

He, Li
McAndrew, Ryan
Barbu, Razvan
et al.

Publication Date

2024-08-01

DOI

10.1016/j.jmb.2024.168650

Peer reviewed

Structure and interactions of HIV-1 gp41 CHR-NHR reverse hairpin constructs reveal molecular determinants of antiviral activity

Li He [1], Ryan McAndrew [3], Razvan Barbu [2], Grant Gifford [2], Cari Halacoglu [2], Camille Allaire [3], Lindsey Weber [3], Line G. Kristensen [3], Sayan Gupta [3], Yan Chen [4], Christopher J. Petzold [4], Marc Allaire [3], Kathy H. Li [5], Corie Y. Ralston [6] and Miriam Gochin [1,5,*]

1 Department of Basic Sciences, Touro University California College of Osteopathic Medicine, 1310 Club Drive, Mare Island, Vallejo, CA 94592, USA

2 Master of Science in Medical Health Sciences, Touro University California College of Osteopathic Medicine, 1310 Club Drive, Mare Island, Vallejo, CA 94592

3 Molecular Biophysics and Integrated Bioimaging Divisions, Lawrence Berkeley National Laboratory, Berkeley, CA 94720

4 Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720

5 Department of Pharmaceutical Chemistry, UCSF School of Pharmacy, San Francisco, CA 94143, USA

6 Molecular Foundry, Lawrence Berkeley National Laboratory, Berkeley, CA 94720

* Correspondence: mgochin@touro.edu; Tel. 1-707-638-5463

ABSTRACT

Engineered reverse hairpin constructs containing a partial C-heptad repeat (CHR) sequence followed by a short loop and full-length N-heptad repeat (NHR) were previously shown to form trimers in solution and to be nanomolar inhibitors of HIV-1 Env mediated fusion. Their target is the in situ gp41 fusion intermediate, and they have similar potency to other previously reported NHR trimers. However, their design implies that the NHR is partially covered by CHR, which would be expected to limit potency. An exposed hydrophobic pocket in the folded structure may be sufficient to confer the observed potency, or they may exist in a partially unfolded state exposing full length NHR. Here we examined their structure by crystallography, CD and fluorescence, establishing that the proteins are folded hairpins both in crystal form and in solution. We examined unfolding in the milieu of the fusion reaction by conducting experiments in the presence of a membrane mimetic solvent and by engineering a disulfide bond into the structure to prevent partial unfolding. We further examined the role of the hydrophobic pocket, using a hairpin-small molecule adduct that occluded the pocket, as confirmed by X-ray footprinting. The results demonstrated that the NHR region nominally covered by CHR in the engineered constructs and the hydrophobic pocket region that is exposed by design were both essential for nanomolar potency and that interaction with membrane is likely to play a role in promoting the required inhibitor structure. The design concepts can be applied to other Class 1 viral fusion proteins.

INTRODUCTION

Viral infection by HIV-1 relies on the fusion of cellular and viral membranes mediated by the transmembrane glycoprotein-41 (gp41). The extracellular domain of gp41 consists of N-heptad repeat (NHR) and C-heptad repeat (CHR) domains which undergo conformational rearrangement into a trimeric hairpin structure, also known as a 6-helix bundle (6-HB), during fusion [1-4]. Fusion can be prevented by interrupting the conformational change of gp41[5]. Both peptide and small molecule inhibitors of gp41 have been explored as potential inhibitors of viral fusion. Small molecules have been difficult to develop, both because of the requirement to disrupt a protein – protein interface, and by difficulties in experimental confirmation of ligand binding [6]. C-peptides such as Enfuvirtide® (T-20) and C34 are nM fusion inhibitors that target the pre-hairpin in situ NHR intermediate [7, 8]. Major contributors to their potency are residues from the membrane proximal external domain (MPER) in T-20, or residues that interact in a known deep hydrophobic pocket (HP) in C34. N-peptides are μM fusion inhibitors, due to their tendency to aggregate [9]. They may target in situ CHR as trimers, preventing formation of the 6-HB, or in situ NHR as monomers or dimers, preventing endogenous NHR trimer formation [10, 11]. Stabilization of N-peptide trimers has been associated with vastly increased potency but limits the mode of action to the CHR target[11]. N-helical trimer stabilization has included single chain protein mimetics [12, 13] , NHR peptides fused to a soluble trimerization domain [9, 14], addition of interhelical disulfide bonds [15-17], and mutation of residues involved in the trimerization domain[18]. Many of the resulting constructs displayed nM to sub-nM activity against viral fusion.

Enfuvirtide® is the only approved peptide fusion inhibitor targeting gp41, and unfortunately its use leads to rapid development of resistant viral strains [19, 20]. On the other hand, viral resistance to N-helical mimics develops more slowly [21, 22]. Resistance takes the form of

increased endogenous 6-HB stability [11, 23], enabling the virus to compete better against the inhibitors. Unlike C-peptides, where potency is correlated to binding affinity, N-peptide potency is strongly dependent on association kinetics [24, 25], implying a limit to potency for larger NHR mimetics, and a decoupling of potency and binding affinity [26]. Stabilized trimers are also less flexible in dealing with resistance [11]. Intriguingly, studies showed that gp41 CHR was accessible to an N-helical trimer mimetic in native Env, especially after CD4 binding [5], while gp41 NHR was not available to C-peptide inhibitors until after co-receptor binding [24]. These results were supported by a later cryo-EM structure of the metastable prefusion Env, which showed a break in the NHR helix at the HP region while the C-terminal part of the CHR was a well-formed helix [27].

More recently, Jurado et al, using a single polypeptide mimetic CovNHR, dissected the NHR-CHR energy landscape, finding that multiple cooperative interactions between adjacent sub-pockets on the NHR trimer were required for high binding affinity[28]. They followed up with an antiviral study of miniproteins "CovNHR-N" and "CovNHR-C", which each represented half of the total NHR [29]. CovNHR-C included the HP. They found that a disulfide-stabilized form of CovNHR-N was a low to mid-nM inhibitor against several viral strains, while CovNHR-C, which was highly stable, was inactive. Both bound with μM affinity to their respective C-peptides. Taking into account the kinetic dependence of fusion inhibition, they reasoned that the potency of CovNHR-N was due to low kinetic restriction and access to CHR in the prefusion intermediate. The authors went on to suggest a way forward by optimizing the smaller CovNHR-N construct. However, CovNHR-N miniproteins were still an order of magnitude less potent than CovNHR. Given the finding of long range allosteric effects throughout the NHR structure[28], separation into two miniproteins may not adequately reflect the contribution of each component. The inactivity of CovNHR-C alone may be due to steric or kinetic hindrance or

aggregation, as it covers a region of gp41 that is well-established as playing a critical role in the NHR-CHR interaction and its inhibition [30-32].

Reverse hairpin constructs of gp41 ectodomain prepared by our group provide an opportunity to further explore the contribution of different parts of the NHR to antiviral activity, without separating them into separate entities. They contain a small C-helical segment and long NHR, with the CHR residues preceding the NHR residues in the primary sequence[33]. When arranged in a "reverse hairpin" structure (Figure 1), the highly conserved hydrophobic pocket (HP), a target for fusion inhibitors, is exposed, while the remainder of the NHR is covered by CHR. There are no covalent structure stabilizing modifications such as disulfide bonds. This construct was initially designed as a receptor for examining low molecular weight inhibitors of HIV fusion, envisioning that CHR coverage could ameliorate aggregation that has previously limited the use of NHR peptides in biophysical studies or as fusion inhibitors. Biophysical studies, including CD, NMR and AUC, revealed that the molecules formed unique trimers, with high helical content, and that they bound to HP binding peptides and small molecules [33]. Unfolding occurred with elevated temperatures in the presence of membrane mimetic DPC [34]. The regions covered by CovNHR-N and CovNHR-C [29] are

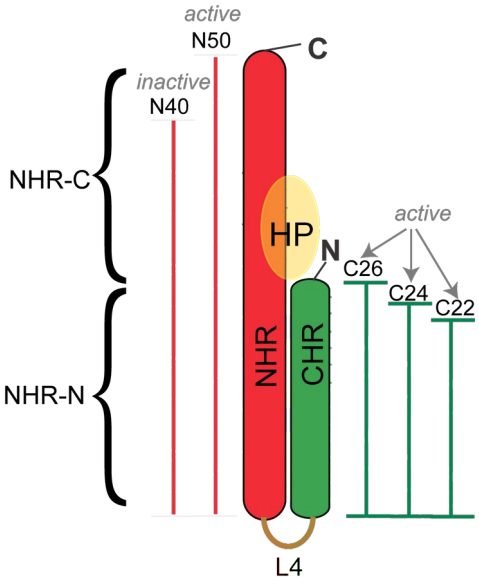


Figure 1. Schematic representation of the reverse hairpin constructs, with truncated CHR domain in green and the NHR domain in red. N- and C-termini are labeled. The position of the HP is shown in yellow, noting that it is present only in the trimeric structure (monomer shown). N- and C-terminal halves of the NHR helix are indicated as NHR-N and NHR-C. Active and inactive truncations are also indicated.

indicated in Figure 1 using NHR-N and NHR-C for reference. In the folded hairpin, NHR-N is covered by the CHR domain.

Interestingly, the reverse hairpins were found to be nM inhibitors of HIV fusion [34], as was observed for other NHR-trimer mimetics. This might be attributed to the exposed HP, a critical hotspot in the fusion reaction, since a hairpin with covered HP had no antiviral activity. We observed that N-terminal truncation (of 8 CHR residues) did not alter antiviral activity, while C-terminal truncation (of 5 or 10 NHR residues) rapidly led to loss of antiviral activity (Figure 1). Exposed HP residues were present in all constructs, but trimer stability was severely impacted by C-terminal truncation. We concluded that the inhibitor was active as a trimer.

The insensitivity of antiviral potency to CHR length led us to ask whether only NHR residues are important in interacting with in situ gp41. Since the hairpin inhibitors nominally display only a partially exposed N-helix, their exact mechanism of action remains unclear. If the CHR were absent altogether, μM inhibitory activity would be expected [18], implying that even truncated CHR must play a role. Here we have conducted X-ray crystallography studies of the designed hairpin and of a next generation hairpin with altered CHR sequence. Hairpin folding in solution was examined using circular dichroism (CD) and fluorescence. We also examined the effect of hairpin stabilizing mutations or HP blocking on antiviral efficacy. We concluded that a fully exposed NHR must be available in the milieu of the fusion environment, and that the antiviral efficacy does not reside solely on the availability of HP residues. However, the HP residues play a critical role, conveying a third to a half of the antiviral potency of the full length NHR. CHR, meanwhile, may interact with the membrane surface, exposing the NHR and drawing the inhibitor to the site of the fusion reaction.

RESULTS

1. Hairpin Design and Modifications

We originally designed a reverse hairpin named C28(L4)N50 (Figure 2), in which 28 residues of the CHR (637-664, HXB2 numbering) were followed by a 4 residue loop (SGGG) and 50 residues of the NHR (542-591, HXB2 numbering) [33]. An N-terminal proline was added in the construct, as part of an acid labile Asp-Pro sequence included to enable formic acid cleavage after high level expression and His-tag purification. Two residues (KR) were added at the C-terminus. This protein had a calculated pI of 9.34, was soluble at acidic pH (≤ 5.5) and relatively insoluble at neutral pH (pH 7.4). As part of the exploration to obtain structural details of the interaction of small molecules in the HP, we designed a second generation reversed hairpin C26'(L4)N50, as well as versions with shorter C-segments C24'(L4)N50 and C22'(L4)N50. The apostrophe indicates modification in the sequence of the C-segment compared to the original constructs, to include a series of E_iE_{i+1}-$R_{i+4}R_{i+5}$ substitutions which have the potential to form salt-bridges on the CHR interface in solution (Figure 2). These are predicted to improve solubility and helix stability [35], while leaving the CHR-NHR interaction untouched. No changes were made to the NHR sequence except for K78 which was mutated to arginine in some preparations; this mutation had no effect on the protein properties or behavior. Additionally, a change of -4 in the charge of the protein was engineered by exchanging the two non-gp41 residues KR at the C-terminus with ED. The predicted Pi of the second generation constructs was 5.85 and they proved to be soluble at pH 7.4 and insoluble at pH 5.5. Solubility at neutral pH was considered essential for studying small molecule inhibitors at high concentration, since all of our inhibitors contained carboxylate groups that would be protonated at low pH, reducing solubility.

The progressive shortening of the C-helix in C24'(L4)N50 and C22'(L4)N50 mirrored earlier efforts to monitor the effect of CHR length on antiviral potency [33, 34] but was also the result of observations made in the crystal structure of C28(L4)N50 (see below) in which it was apparent that the N-terminal end could potentially block access to part of the HP. Additionally, early crystal structures of C28(L4)N50 and C26'(L4)N50 revealed pairs of residues involved in the

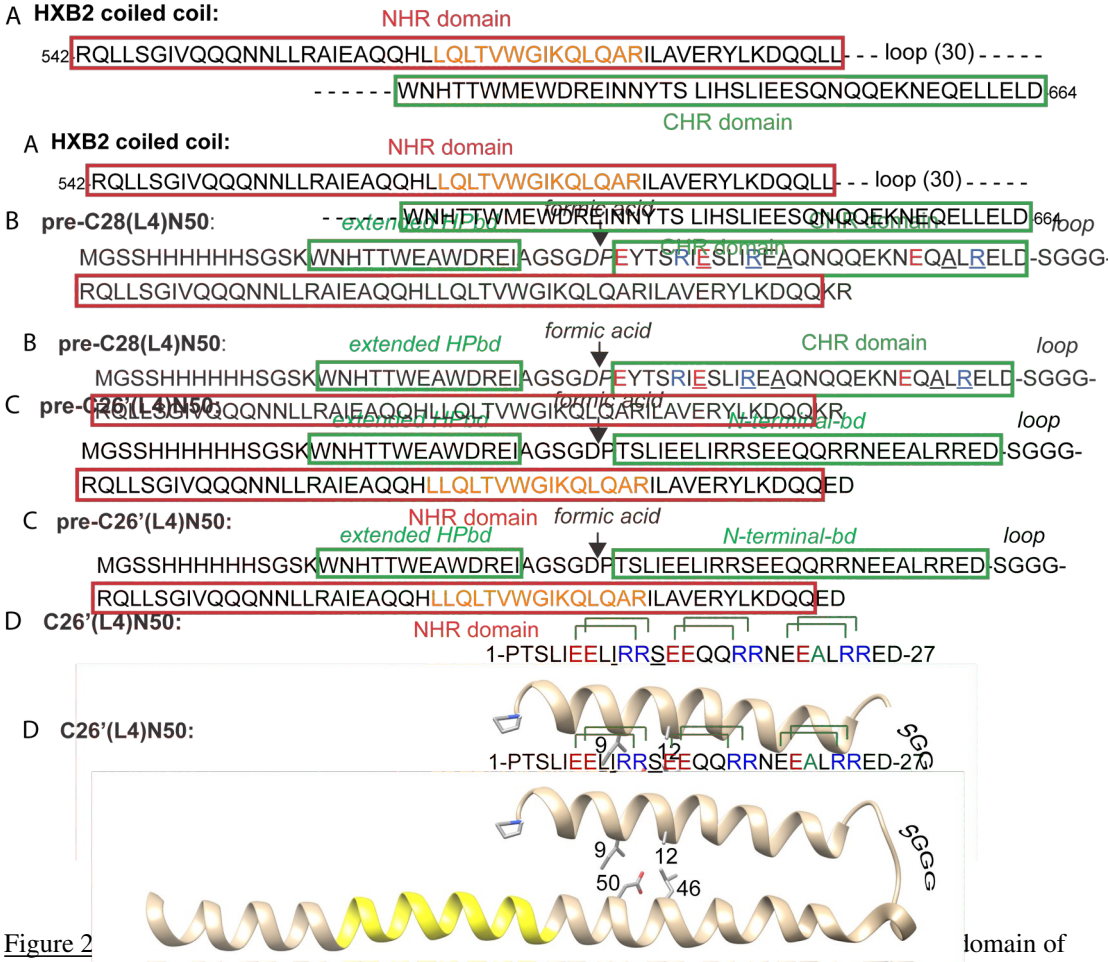


Figure 2. Design of reverse hairpins. A. Sequence of the NHR and CHR coiled coil ectodomain of HXB2 gp41. B. Sequence of the first generation construct pre-C28(L4)N50. C. Sequence of the second generation construct pre-C26'(L4)N50. B and C show the site of formic acid cleavage used to obtain the final constructs. *bd* stands for binding domain. D. Alignment of the sequence of C26'(L4)N50 on a representation of the hairpin structure observed by crystallography. The Glu-Arg modifications in this sequence are shown in red and blue together with the potential salt bridges. Residues ED at the C-terminus are also highlighted; these were KR in the first generation reverse hairpins. Residues that were substituted in pairs for cysteine modification are underlined and illustrated in the depicted structure. The region encompassing the HP is in yellow. The C24' and C22' sequences are identical to that of C26', except that they lack two or four residues, respectively, at their N-termini, hence beginning "PLIEELI" or "PEELI", respectively.

CHR – NHR interaction that could be substituted with cysteines to potentially form disulfide bonds in the folded hairpin. Two pairs that were deduced from study of the crystal structure are shown in Figure 2 and were prepared as C26'(L4)N50-C_9C_{50} and C26'(L4)N50-$C_{12}C_{46}$. An additional cysteine containing construct with the original C-helix sequence C26(L4)N50-C_9C_{50} was also prepared. These proteins proved useful for elucidating the contribution of NHR-N to antiviral activity.

2. Crystal structure of the hairpins C28(L4)N50, C26'(L4)N50 and C24'(L4)N50

Proteins were crystallized from 10mg/ml stocks in sodium formate using the hanging drop technique. Crystallization conditions were obtained after conducting crystal robot screening (See Materials and Methods). The proteins proved amenable to crystallization, forming crystals up to 60 μM in diameter. Crystals of C28(L4)N50 formed in Hampton Custom Reagent #20 containing 200 mM ammonium sulfate, 100 mM sodium acetate either at pH 4.2 and 15% w/v PEG-4000, or at pH 4.6 and 25% PEG 4000. Crystals of C26'(L4)N50 and C24'(L4)N50 formed in 0.1M lithium sulfate, 0.1M sodium citrate trihydrate at pH 6.4 and 25% PEG-1500. Data and statistics from the crystallographic experiments are provided in Table S1 (Supplementary Data).

Small molecule HP inhibitors were introduced into many of the crystal preparations in order to explore protein - ligand binding, either by co-crystallization or by soaking. Some of the small molecules contained an activated STP ester, enabling covalent association with the hairpin via a hydrophobic pocket lysine [36, 37]. We were not able to find crystallization conditions for samples in which ligands were added to the hanging drop or when the protein – ligand covalent adducts were preformed and purified. However, crystals of the various hairpins were resilient to

soaking for hours to days in solutions containing ligands. Ligands used are shown in Figure S2 (Supplementary Data). Crystal structures were obtained for C26'(L4)N50 without ligand as well as after soaking in compound **2** or after attempted co-crystallization with **4.** A crystal structure of C26'(K4)N50-K78R was solved after soaking in **2-STP**, The structure of C24'(L4)N50-K78R was solved after soaking for 24 hours with either **1-STP** or **2-STP**. Structures were also obtained for C28(L4)N50 without ligand and after soaking with **3**. Details of ligand concentrations and soak conditions are provided in the Supplementary Data, Tables S2A and S2B.

Five crystal structures were refined and are reported in the PDB. Details of the refinements are provided in the Supplementary data Table S1. The crystal structure of the hairpin trimer C26'(L4)N50-K78R is shown in Figure 3A. With symmetry applied the protein is trimeric, as predicted by design and observed in solution experiments. The L4 loop plus three residues before and one residue after the loop were not defined. Two Arg residues near the C-terminus, R75 and R78, coordinated a SO_4^{2-} ion, stabilizing the crystal structure.

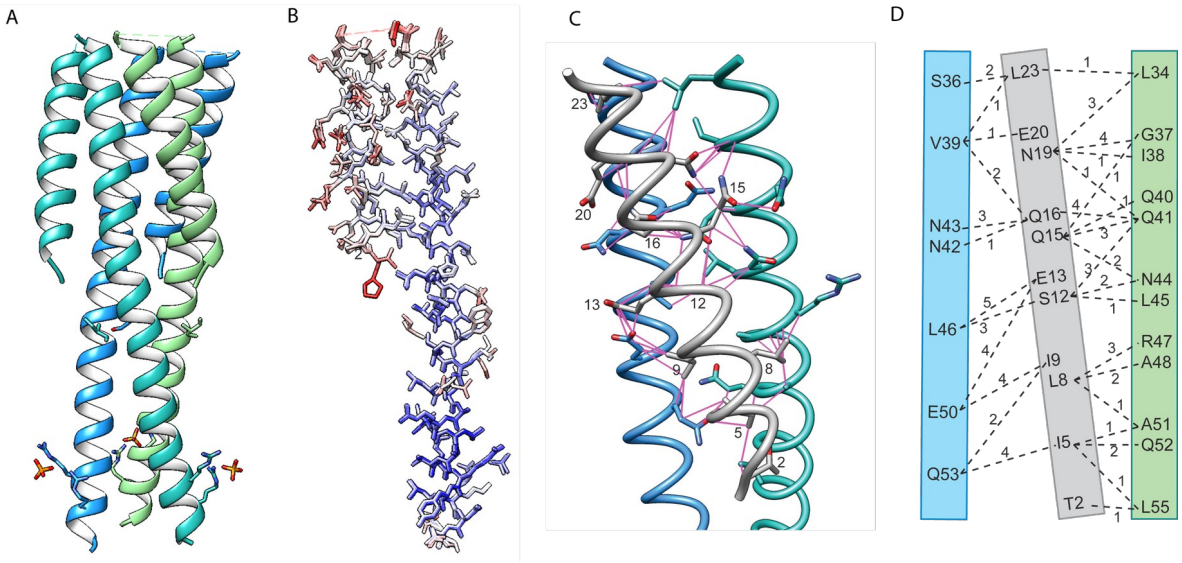
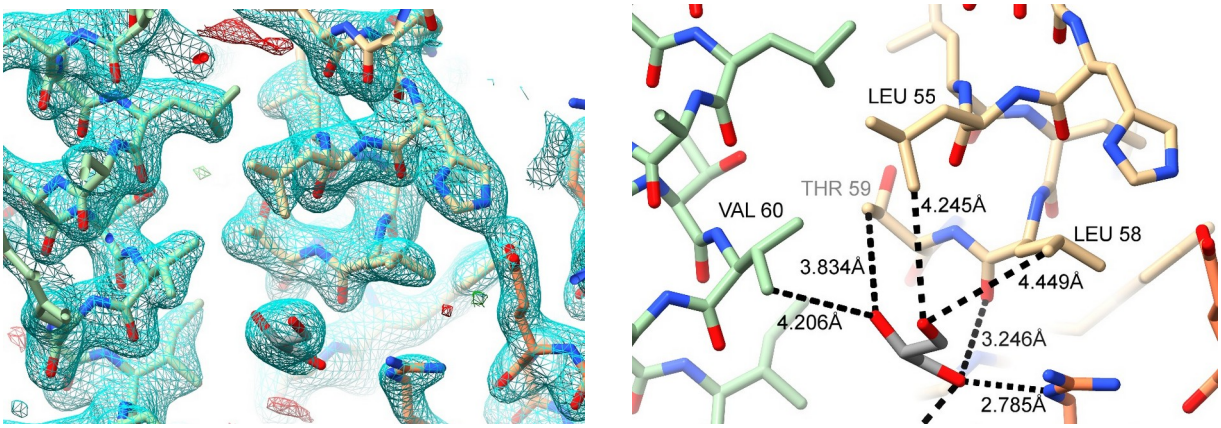


Figure 3. Crystal structure of reverse hairpins. A. Trimeric hairpin C26'(L4)N50. B. Overlay of hairpin monomers of three C26'(L4)N50 structures and one C24'(L4)N50 structure colored according to B-factors (blue – red). C. Contacts (in magenta) between one CHR helix (gray) within the groove formed by two NHR helices (blue and green). D. Schematic representation of C. Residues that make CHR-NHR interhelical contacts are connected by dashed lines, with the number of contacts shown.

Four separately obtained crystal structures of the monomeric hairpin are shown overlaid in Figure 3B, with the atoms color coded according to the B-factor, which ranged from 17 (deep blue) to 103 (deep red). The crystals in these overlaid structures were soaked with different ligands or under different conditions (Supplementary Data S2). Main chain atoms and most side chain orientations were superimposable, even between C26'(L4)N50 and C24'(L4)N50, and little to no electron density was found in the HP. In two of the structures, residual electron density found in the HP could be modeled as a glycerol molecule. However, no other density corresponding to the full ligand was found. The B-factors showed that the long (NHR) helices were well defined, while there was considerable flexibility on the short arm of the hairpin (CHR). Side chains pointing into the interior of the trimer interface, or between NHR and CHR were better defined by the electron density than side chains pointing towards solvent. There was no evidence of Glu-Arg salt bridges on the exterior of the CHR-helix, associated with the $E_iE_{i+1}R_{i+3}R_{i+4}$ sequence built into the structure with the hope of stabilizing the helix [35]. The Lys 64 side chain (Lys 62 in C24'(L4)N50), which is the target of covalent or H-bond interactions with ligands, was not well-defined, with conformational variability in the four structures. Analysis using Chimera [38] revealed 69 contacts between the CHR helix (C26') and two NHR helices which formed the groove into which C26' was docked (Figure 3C, D). Most of these are non-polar or van der Waal's contacts (Supplementary Data Table S3), but included 8 hydrogen bond interactions.



Crystal packing contacts included pairs of intermolecular salt bridges at the end of the NHR helix, involving residues R75 and D79 on neighboring molecules (residue numbering for C26'(L4)N50). Crystal packing also provided an explanation for the lack of observable ligand density in the HP, apart from the residual electron density that could be accommodated by a small additive. Figure 4 shows the HP region and residues in a helical segment of a symmetry-related neighboring molecule (shown in orange). There were several potential hydrogen bonds to HP

Table 1. Intermolecular H-bond interactions between HP residues and a neighboring molecule in the unit cell

HP residue / atom †	**Residue / atom of symmetry-related neighboring molecule †**	**Distance**
His54 (52) side chain $N\varepsilon_2$	Glu20 (18) side chain carboxylate	2.49Å
Trp61(59) indole $N\varepsilon_1$	Glu21 (19) side chain carboxylate	3.30Å
Lys64 (62) side chain εNH_2	Glu21 (19) side chain carboxylate	3.40Å §
Gln 65 (63) side chain O	Arg24 (22) side chain CB	3.26Å
Gln 65 (63) side chain N	Arg 24 (22) main chain O	3.20
Glycerol O	Arg24 (22) side chain Nε	2.78Å

† residue numbering for C26'(L4)N50, with that of C24'(L4)N50 in brackets. § Tentative, Lys side chain not well-defined by electron density.

residues H54, W61 and Q65 that stabilized the intermolecular interaction with the helical segment (Table 1). Residue R24 near the loop (R22 in C24'(L4)N50) in the neighboring molecule made several contacts, including with the Q65 side chain and with a glycerol molecule that was modeled to fit the extra electron density in the HP. The distance between the glycerol -OH and R24 NH was 2.78 Å. K64 is highly flexible in the structure, and may or may not participate in intermolecular H-bond interactions.

Virtual docking simulations with the ligands used in soaking or co-crystallization experiments yielded docked poses that would clash with the protruding R24 side chain, preventing ligand approach to the pocket. Despite screening several common crystal conditions, no crystals were formed with pre-bound covalent ligand – protein adducts. Protrusion of R24 from the symmetry related neighboring molecule would not only impede a ligand or a K64-ligand adduct, but potentially causes K64 to face away from the ligand binding site. It is likely that the crystal packing requirements prevented formation of crystals of protein – ligand complexes.

3. Hairpin structure in solution *3.1 CD studies indicated trimer in solution* The crystal structure indicated a well-folded trimeric hairpin, although the loop was disordered. CD experiments confirmed the coiled coil structure in solution, with high helical content and the ratio $\theta_{222}/\theta_{208} > 1$ (Table 2 and Figure 5). Table 2 reports the % helicity, melting temperature and

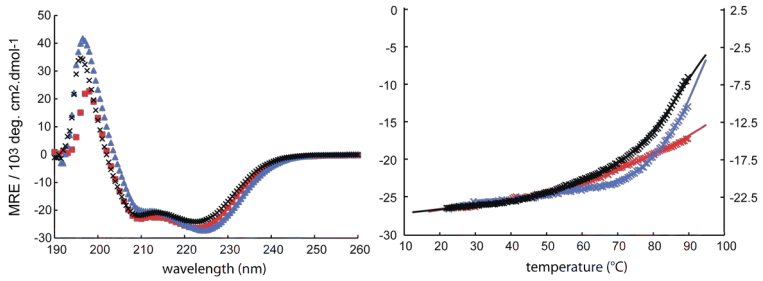


Figure 5. CD spectra and temperature melts for C26'(L4)N50 (red), C24'(L4)N50 (blue) and C22'(L4)N50 (black). Experiments were conducted in PBS. For ease of visualization, the secondary y-axis in the melting curve plot is slightly shifted to align the melting curve of C22'(L4)N50 with the others. The solid lines show the fit to the data to obtain the melting temperature (see text).

$\theta_{222}/\theta_{208}$ ratio for each protein. The $\theta_{222}/\theta_{208}$ ratio is sensitive to the oligomerization state of helices, with helices that are arranged in a coiled coil structure typically having a value > 1[39]. Melting curves were obtained by measuring θ_{222} as a function of temperature over the range 22 – 90°C. They were analyzed by fitting to a two-state trimer to monomer transition with five parameters, the molar ellipticities for the folded and unfolded state, θ_F^{MRE} and θ_U^{MRE}, the melting temperature T_M at the midpoint of the unfolding transition, the enthalpy of unfolding ΔH_F and the change in heat capacity, ΔCp (see Materials and Methods) [40, 41]. The results of the fits are reported in Table 2. The values obtained for θ_U^{MRE} were consistent with unfolded protein. The CD data proved insensitive to the ΔCp values (< 1 kcal/mol.K), which were consequently not reported. The melting temperature T_M of the main structural transition was dependent on the length of the CHR segment, dropping from > 100° to 90° to 85°C for CHR lengths of 26, 24 and 22 residues, respectively. Clearly, all of the constructs were highly stable in PBS. Due to minimal melting of C26'(L4)N50 up to 90°C, only a lower limit for the T_M could be obtained. All melts were reversible, with CD spectra returning to their pre-melt form. Along with the reduction in T_M, the calculated enthalpy of the transition also decreased with CHR length, with values indicative of a stable folded state [42]. The results

Table 2: CD studies of reverse hairpins in PBS

	Observed *			**Calculated**			
Protein	**θ_{MRE} / 10^3 ¶**	**% helix ¶**	**θ_{222} / θ_{208}**	**T_M / °C §**	**θ_F /10^3 [a]**	**θ_U /10^3 [b]**	**ΔH_{fold} / kcal.mol^{-1} [b]**
C26'(L4)N50	-26.5	80	1.2	> 100	n.d.	n.d	n.d.
C24'(L4)N50	-26.4	81	1.1	90	-27.0	-6.0	100
C22'(L4)N50	-24.1	73	1.0	85	-26.5	-5.0	60

* 20μM protein solutions measured at 22°C; ¶ mean residue ellipticity at 222 nm, accuracy ± 4%, assuming 100% helix at MRE_{222} = -33*10^3 deg.cm^2.dmol$^{-1;}$ §midpoint of unfolding transition; [a]MRE_{222} of folded state; [b]free energy of unfolding; T_M, θ_F, θ_U and ΔH_{fold} were determined by fitting the data to equation 3. Error range ± 0.4 x 10^3 deg.cm^2.dmol^{-1}for θ_F and θ_U, ± 2 kcal/mol for ΔH_{fold}; n.d. not determined due to insufficient sampling of the melting curve

indicated that CHR-NHR interactions contributed to trimer stability and played a role in the energy of unfolding.

3.2 Fluorescence studies confirmed the hairpin structure in solution The hairpins contain a single Trp residue, residing in the HP, and its proximity to the N-terminus of the folded hairpin (Figure 1) could be examined through Fluorescence Resonance Energy Transfer (FRET) to an acceptor attached at the N-terminus. Trp (ex 280nm, em 350nm) and IAEDANS (5-((((2-Iodoacetyl)amino)ethyl)amino)Naphthalene-1-Sulfonic Acid) (ex 335nm, em 455-465nm) were used as an established donor - acceptor pair. IAEDANS was attached to the cysteine residue of an S5C mutant of C28(L4)N50. Fluorescence of the single Trp residue in the HP was followed by excitation of a 1.5 μM solution in both unlabeled and IAEDANS-labeled C28(L4)N50. Strong emission was observed at 350nm and 465nm, respectively, from which it was possible to deduce an approximate distance between Trp and IAEDANS of ~12 Å, a distance that could be satisfactorily modeled from the crystal structure of the folded hairpin. Example spectra and details of the calculation are provided in the Supplementary Data (S4).

4. Hairpin structure in solution in the context of the viral fusion reaction

4.1 CD studies showed reduced hairpin stability and destabilized trimer in membrane mimetic solvent. The observation of low nM antiviral potency suggested that, in the environment of the cells, the CHR segment may disengage in the reverse hairpin structures, exposing full length trimeric NHR, as was required in other N-trimer constructs with similar antiviral potency. This hypothesis was bolstered by the observation of higher disorder for the CHR segment in the crystal structure. To test this, we repeated the CD experiments in the presence of a membrane mimetic solvent dodecylphosphocholine

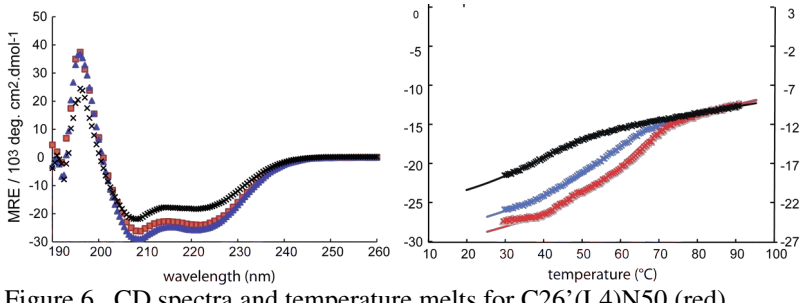


Figure 6. CD spectra and temperature melts for C26'(L4)N50 (red)

(DPC). DPC is a zwitterionic detergent that has been used in biophysical studies of membrane-associated peptides and proteins, due to its non-interfering spectroscopic properties[43]. It forms micelles with a CMC ~ 1mM [44]. Although DPC micelles are a poor representation of the microenvironment around in situ gp41 [45-50], they are useful in this biophysical study of folding because they destabilize the hairpin structure allowing for transitions to be observed at a more amenable temperature[34]. This was borne out in CD results (Figure 6 and Table 3) which showed weakening of the coiled coil structure. Thermal stability was reduced by more than 40°C for each protein, with a strong correlation to the length of the CHR segment. The T_M decreased from 56° to 48 to 30°C for CHR lengths of 26, 24 and 22 residues, respectively, when DPC was present. The lower stability of the coiled coil structure in DPC was reflected in the accompanying reduction in the enthalpy of the melting transition, both with reduction in CHR length and when compared to the same protein in the absence of DPC. Thus CHR opening is likely to be on the unfolding pathway for the hairpins. Fitting of the melting data to the trimer – monomer transition model depended only on the difference between θ_F^{MRE} and θ_U^{MRE}, and not on their absolute values, so we chose θ_F^{MRE} in DPC to be the same as in buffer. Interestingly, the fitted θ_U^{MRE} were much more negative than in buffer, indicating residual helicity after the transition that inversely correlated to CHR length: 57%, 61% and 64%, respectively, for C26'(L4)N50, C24'(L4)N50 and C22'(L4)N50. Since the values tracked approximately with percentage of the construct that is NHR, they indicated that the partial helical structure retained in the presence of DPC micelles was likely NHR.

Thus DPC affected both NHR-CHR and NHR-NHR interactions, exposing full length NHR and promoting trimer → monomer transitions. The T_M dependence on CHR length implied that CHR interaction with the micelles was a key part of the melting transition. The melts were

reversible with the spectrum returning to its pre-melt form in each case when the temperature was lowered to 22°C. .

Table 3: CD studies of reverse hairpins in PBS and 10mM DPC

	Observed *			Calculated			
Protein	**θ_{MRE} / 10^3 ¶**	**% helix ¶**	**θ_{222} / θ_{208}**	**T_M / °C §**	**θ_F /10^3 [a]**	**θ_U /10^3 [b]**	**ΔH_{fold} / kcal.mol^{-1} [c]**
C26'(L4)N50	-27.5	83	0.9	56	-27.5	-19.0	55
C24'(L4)N50	-25.9	78	0.9	48	-27.0	-20.0	48
C22'(L4)N50	-18.7	57	0.8	30	-26.5	-21.1	40
C26'(L4)N50 -$C_{12}C_{46}$ [d]	-25.1	76	0.85	60	-27.5	-21.5	27

* 20μM protein solutions measured at 22°C; ¶ mean residue ellipticity at 222 nm, accuracy ± 4%, assuming 100% helix at MRE_{222} = -33*10^3 deg.cm^2.dmol$^{-1;}$ §midpoint of unfolding transition; [a]MRE_{222} of folded state; [b]MRE_{222} after transition; [c]free energy of unfolding; T_M, θ_F, θ_U and ΔH_{fold} were determined by fitting the data to equation 3, assuming the same value for θ_F as in buffer; [d]di-cysteine mutant, see text.

4.2 *Fluorescence confirmed interaction with DPC and fraying of the hairpin* Fluorescence studies of hairpins in the presence of DPC corroborated the CD data. Tryptophan fluorescence shifted to lower wavelength and increased in intensity with increasing concentrations of DPC (data not shown), consistent with partitioning of NHR into a more hydrophobic environment[45]. Although no direct comparison of FRET can be made between non-micellar and micellar environments (Supplementary Data S4), heating the sample containing DPC from 25°C to 37°C resulted in a calculated 5% increase in the distance between IAEDANS probe at the N-terminus and the Trp residue in the hydrophobic pocket. This provided evidence that DPC was causing the CHR to disassociate from NHR, even early in the melting transition (T_M = 79°C) [34].

The behavior in DPC micelles could hint at the role of membrane bilayer during fusion, i.e. by effecting dissociation of CHR from the NHR trimer. However, DPC also destabilized the trimer, which is at odds with our previous hypothesis that trimeric NHR coiled coil was the determinant of antiviral potency in the reverse hairpins [34]. One explanation is that DPC is an

imperfect mimic of the membrane, as has been observed previously, but a model in which the reverse hairpins act to disrupt endogenous NHR trimer formation cannot be ruled out.

4.3 *A disulfide bond linking CHR to NHR altered the structural transition* Cysteine pairs were introduced into the constructs in order to potentially form disulfide bonds that could restrict hairpin opening by linking NHR-N and CHR (Figures 1 and 2). Three cysteine-containing proteins were prepared. C26'(L4)N50-C_9C_{50} contained mutations I9C and E50C and C26'(L4)N50-$C_{12}C_{46}$ contained mutations S12C and L46C. An additional mutant C26(L4)N50-C_9C_{50} contained the original CHR sequence plus the two indicated cysteine mutations. Chimera was used for virtual cysteine substitution in the crystal structures and C_α-C_β torsion angles were adjusted to optimize the distance between the two sulfur atoms (expected to be 2.05Å for an ideal disulfide). An S-S distance of 2.05Å could easily be achieved for the $C_{12}C_{46}$ pair with a corresponding S-S torsion angle (X_3) of 90 - 93°, which lies within the range found in native disulfide bonds [51]. The shortest distance achievable for the C_9C_{50} pair was 2.57Å.

A spectrophotometric assay using thiol sensitive Ellman's reagent (5,5'-dithio-bis-[2-nitrobenzoic acid) was used to estimate the percentage of residual free thiols in each sample (Materials and Methods). The results are shown in Table 4. They indicate that cysteine substitutions at positions 12 and 46 successfully led to formation of a disulfide bond connecting the NHR and CHR helices, while cysteines substituted at positions 9 and 50 were not or only partially able to form a disulfide bond. This is in keeping with the predictions for disulfide bond formation from the crystal structures.

Table 4. Determination of free thiols from a spectrophotometric DTNB assay

Protein	percent free thiol[‡]
C26'(L4)N50-$C_{12}C_{46}$	7 ± 6
C26'(L4)N50-C_9C_{50}	42 ± 5
C26(L4)N50-C_9C_{50}	89 ± 7

[‡]3 repeat measurements ± standard deviation. 100% free thiol is equal to twice the protein concentration

The CD spectra of C26'(L4)N50 and C26'(L4)N50-$C_{12}C_{46}$ were similar in both PBS and in PBS+DPC within the limits of experimental accuracy (Figure 7A). The melts revealed differences between the two variants. In PBS, slightly greater retention of helix was obtained over the studied temperature range for the disulfide variant (Figure 7B), and as before parameters for the melting transition could not be obtained due to insufficient sampling. In the presence of DPC, the melting curve of C26'(L4)N50 -$C_{12}C_{46}$ flattened considerably compared to that of C26'(L4)N50 (Figure 7C). The curve could be fit with a slightly higher T_M = 60°C, reduced enthalpy of unfolding, ΔH_{fold} = 27 kcal/mol, about half of the value obtained for C26'(L4)N50, and by including ΔCp = 1 kcal/mol.K. In addition the dynamic range θ_F-θ_U was reduced. The values are reported in Table 3.

While these results might suggest retention of structure consistent with formation of a

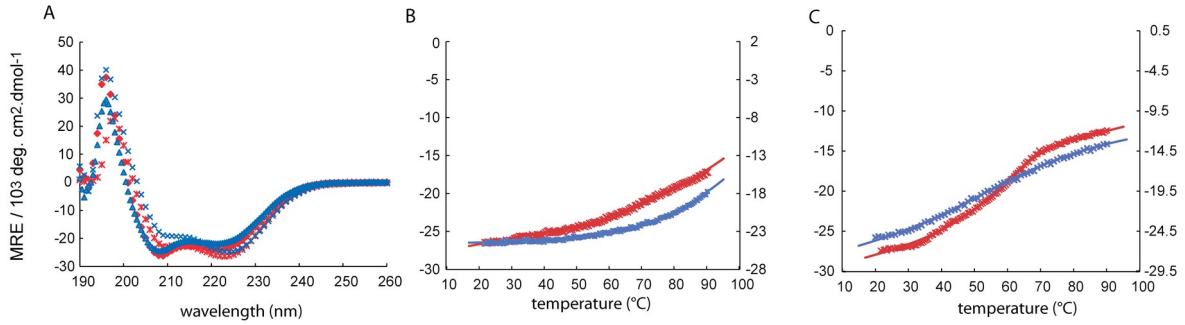


Figure 7. CD spectra of C26'(L4)N40 (red) and C26'(L4)N50-$C_{12}C_{46}$ (blue) in PBS (cross-hatches) and PBS+10 mM DPC (solid symbols). A. Spectra at 22°C. B. and C. Temperature melts in PBS in the absence (B) or presence (C) of 10 mM DPC. Axes on the right in B and C are slightly shifted from those on the left to provide ease of comparison; they refer to the MRE of C26'(L4)N50-$C_{12}C_{46}$.

disulfide bond between CHR and NHR, they were complicated by the observation that a reduced degree of melting in PBS and lack of inflection in the melting curves in DPC carried over to all di-cysteine variants, even those with low percentage dithiols in the native state (Supplementary Data Figure S5). In fact, curve flattening for the other di-cysteine variants was worse, and

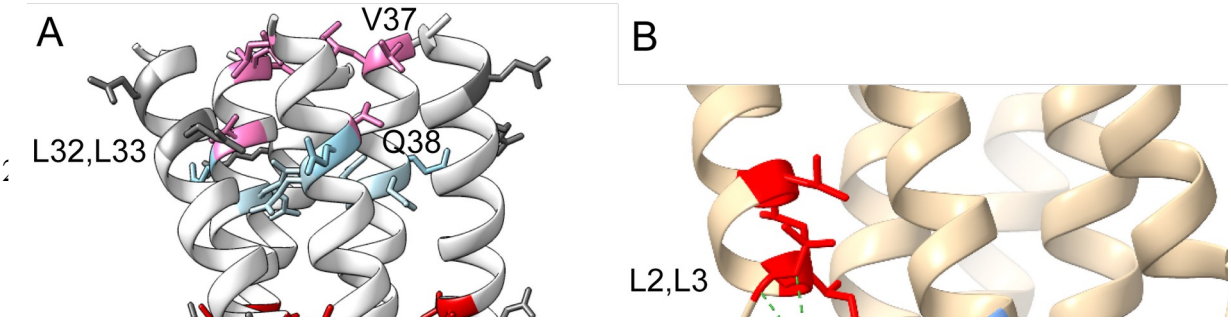
parameter fitting could not be achieved. Reasons could include interference of cysteine side chains in measurement of the helix to coil transition by CD [52], or formation of non-native disulfides at elevated temperature.

5. Hairpin structure in solution with a bound ligand

Given our inability to observe small molecule – protein complexes by crystallography and our interest in parsing different segments of the hairpin for their contribution to antiviral activity, we examined preformed protein – ligand adducts by X-ray footprinting mass spectrometry (XFMS) in solution. XFMS uses changes in solvent accessibility to assess structural differences due to conformational changes and/or ligand binding [53]. Here, we applied XFMS to the solution state C24'(L4)N50 trimer with and without the covalently bound ligand **2**, attached at residue K62. Progressively increasing X-ray dose was used to produce oxidative modifications to certain residues, and quantified using LC-MS/MS, as previously described [54]. The fraction of modification for each residue was fitted to a pseudo first-order rate equation; ratios of the rate constants with and without ligand (Supplementary Figure S7A) show a clear increase in solvent accessibility in the N-terminal region (residues 2-7, gp41 CHR) when ligand is present, and a small increase near the loop (NHR residues 32,33,37). A clear decrease in solvent accessibility of W59, H52 and QQQ(38-40) was observed. The data indicated that ligand attachment increased fraying at the N-terminus of the shorter CHR helix, which crystal structures consistently show is more dynamic and less rigid than the long (NHR) helix according to B-factor analysis (Figure 3). In addition, protection of W59 and H52 in the ligand bound form is suggestive of ligand blocking HP residues and extending towards the N-terminal edge of the

pocket. Solvent accessibility changes are less than one order of magnitude, which are lower than observed in tightly bound protein – ligand complexes [55], but reasonable for a relatively weak binder. We note that the K_D of **2** is ~20 μM[45, 56]; however, a covalent attachment could affect the possible ligand orientations relative to the protein and improve the effective K_D. Changes in NHR residues near the loop (in the region 32-40) are smaller, and include both increased and decreased solvent accessibility. Q38 dominates the mixed modification of Q38/Q39/Q40, based on a qualitative analysis of relative contribution of each glutamine residue to the mixed modification indicated in the LCMS spectra. The changes near the flexible loop could possibly be explained by a minor component of the protein – adduct solution in which S34 is linked to **2**. Although Lys is the predominant target of STP esters [57], Ser is minimally reactive, and a linkage at S34 in the hairpin has been observed by trypsin digest LC-MS/MS upon addition of excess amounts of **2**-STP (data not shown). The Maldi – MS spectrum of the sample used for footprinting was consistent with some Ser labeling, observed as a small amount of doubly-labeled protein (Supplementary Figure S6B).

The observed solvent accessibility changes (See Supplementary Data S7) are mapped onto the structure of C24'(L4)N50 in Figure 8A. A covalent docking simulation of ligand **2** attached to the side chain of K62 was performed in an attempt to identify ligand pose(s) that could account for the observed changes in and around the HP. The flexible side chain covalent docking protocol of AutoDock4 was applied [58, 59], allowing for rotation of the flexible lysine residue during the simulation, but keeping all other residues fixed. The HP is known to be highly malleable, conforming to local environment and interactions [6]. The current crystal structure of



more commonly observed Trp rotamer (PDB structure 2R5B used here) was superimposed onto the structure of C24'(L4)N50, yielding substantial contacts to the ligand, in accordance with experimental observations. The model predicted additional ligand contacts to L56 and Q65, but no experimental data was available for those residues. The increased solvent accessibility of residues 2-7 could be explained by slight displacement near the flexible N-terminus end to accommodate the ligand.

Importantly, the data suggested that the covalently bound ligand had a localized effect on the protein. It blocked the pocket and partially affected NHR interactions with the N-terminus of the short CHR helix. Complete disruption of the hairpin is unlikely as it would be expected to result in dramatic changes to solvent accessibility, which were not seen.

6. Fusion inhibitory properties of the hairpins

The results of antiviral and cell-cell fusion assays are shown in Table 5 and Figure 9. The change in charge of -4 conferred by changing the end of the sequence from KR to DD or ED

Table 5: EC_{50} values, in nM, in antiviral and cell-cell fusion assays

Protein Name	Antiviral activity (HXB2)	Inhibition of cell-cell fusion
Original CHR sequence		
C28(L4)N50 [a]	10 ± 6	10 ± 2
C21(L4)N50 [a]	8 ± 2	10 ± 3
C28(L4)N50-DD [§]	48 ± 3	36 ± 6
New CHR sequence plus C-terminus KR → ED		
C26'(L4)N50	41 ± 6	57 ± 19
C24'(L4)N50	36 ± 7	44 ± 6
C22'(L4)N50	50 ± 8	56 ± 9
Dithiol mutations on NHR and CHR		
C26(L4)N50-C_9C_{50} [b]	54 ± 10	36 ± 5
C26'(L4)N50-$C_{12}C_{46}$ [c]	2000 ± 1000	603 ± 128
C26'(L4)N50-C_9C_{50} [d]	204 ± 45	46 ± 29
Small molecule covalently attached to HP lysine residue		
C26'(L4)N50-εNH-**2** [e]	6500 ± 400	373 ± 178
C24'(L4)N50-εNH-**2** [e]	> 400	n.d.
C22'(L4)N50-εNH-**2** [e]	623 ± 447	284 ± 53

[a] Published in [33]; [§]C-terminus changed from KR to DD; [b] disulfide not indicated by DTNB assay; [c] disulfide indicated by DTNB assay; [d] partial disulfide indicated by DTNB assay; [e] amide bond between lysine-574 εNH_2 and ligand **2** with loss of a H2O molecule

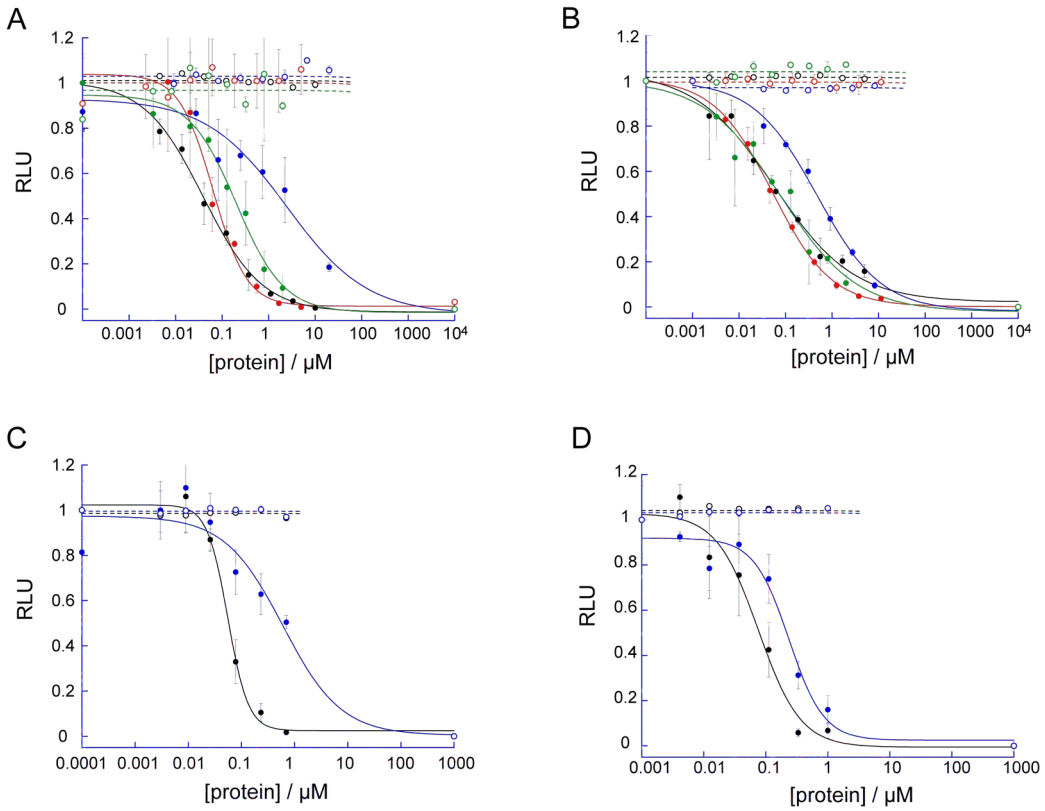


Figure 9. Viral entry (A, C) and cell-cell fusion (B, D) inhibition by protein constructs. A, B: Effect of disulfide bond formation. C26'(L4)N50 (black) and double cysteine mutants C26'(L4)N50-$C_{12}C_{46}$ (blue), C26'(L4)N50-C_9C_{50} (green), C26(L4)N50-C_9C_{50} (red). C, D: Effect of inhibitor bound in the HP. C22'(L4)N50 (black), C22'(L4)N50-εNHCO-**2** (blue). Open symbols indicate cell viability measured by a resazurin assay.

led to a 4-fold increase in the EC_{50}, which may reflect electrostatic repulsion of the inhibitor near the negatively charged membrane surface. No additional effects were noted for inhibitors with the new CHR sequence that also contained the charge change. Experiments measuring virus - cell fusion and those measuring cell – cell fusion gave similar outcomes. The major differences occurred with modifications that resulted either in a disulfide clasp preventing unfolding of the hairpin, or that covered the HP.

The effect of a fully formed disulfide bond between NHR and CHR domains in the hairpin C26'(L4)N50-$C_{12}C_{46}$ was to reduce antiviral potency by a factor of 50. Inhibition of cell – cell fusion was reduced by a factor of 10 – 15 for this construct. We have previously observed

reduced sensitivity for NHR targeting inhibitors in the CCF assay compared to the antiviral assay [36]. This may be related to a gradual loss of Env plasmid in passaged HL2/3 cells in the CCF assay, or to differences in prefusion forms on the cell vs. virus surface. On the other hand, C26(L4)N50-C_9C_{50}, which does not form a disulfide bond, retained the same antiviral potency as C26(L4)N50. The potency change correlated with the extent of disulfide bond formation measured in the sample, strongly suggesting that the N-terminal part of the NHR (NHR-N), outside of the HP, is a key contributor to the antiviral potency, as has been observed previously [29]. To study the contribution of the HP region to potency, we examined adducts between a small molecule **2** and hairpins C22'(L4)N50, C24'(L4)N50 and C26'(L4)N50. The adducts were made by covalent bond formation between an activated carboxylate on ligand **2**, itself a 25 μM binder, with the ε-NH_2 of the lysine residue in the HP[36] (See Materials and Methods). XFMS and docking simulations were consistent with the molecule interacting with key residues in the pocket. The loss of potency of the adducts against cell – cell fusion was of the same order of magnitude as that observed with a disulfide clamp, i.e. going from double digit nM to triple digit nM in cell-cell fusion and up to several μM in virus – cell fusion. Thus, in the context of the full length NHR, HP interactions with in situ gp41 were found to be significant.

DISCUSSION

In this study, we examined the solution and crystal structure and properties of reverse hairpin constructs (CHR-loop-NHR) representing the ectodomain of gp41. These proteins are nM inhibitors of HIV fusion. Results confirmed the structure anticipated by the design, in which truncated CHR enabled access to the hydrophobic pocket on the NHR, a key target of fusion

inhibitors. The crystal structures showed well-formed trimeric 6-helix bundle, both for first generation proteins previously described [33], as well as for second generation proteins with external salt-bridges built into the CHR sequence. A potential for partial interference of N-terminal residues in the HP was observed in the crystal structures of C28(L4)N50 and C26'(L4)N50, which was tested by truncating the CHR to 24 or 22 residues. No change in antiviral activity was observed by CHR truncation. Mobility of N-terminal (i.e. CHR) residues, confirmed in the crystal structure, would likely minimize interference in the HP in solution, a result that concurred with previous observations of HP accessibility to small molecules and peptides, even in the construct containing a 28-residue CHR [33]. The crystal structure did however show evidence of occlusion of the HP by crystal packing interactions involving an arginine residue on a neighboring molecule in the unit cell. This may explain our inability to obtain crystal structures of protein ligand complexes, either by co-crystallization or soaking experiments. Solution structure mirrored the crystal structure, with evidence of well-formed trimeric coiled coil in PBS, as observed by CD and fluorescence studies. The E_iE_{i+1}-$R_{i+3}R_{i+4}$ mutations build into the CHR of second generation proteins to form potential salt bridges were not observed to have a significant effect on helicity, stability or antiviral activity of the hairpins, and E_i to R_{i+3} salt bridges were not detected in the crystal structures.

Given the expectation that full length NHR must be available in the context of the fusion reaction to get nM antiviral activity, we examined the dynamics of protein unfolding by CD and fluorescence experiments. CD melts in PBS demonstrated a highly stable trimeric core structure. A slight decrease in T_M and notable decrease in ΔH_F accompanied the decrease in CHR length from 24 – 22 residues, indicating that CHR interactions helped to stabilized NHR trimer. We also conducted experiments in the presence of 10 mM zwitterionic dodecyl-phosphocholine

detergent DPC, reasoning that inhibitor interaction with the membrane may play a role in its fusion inhibitory activity by exposing full length NHR. In the presence of DPC, significant changes occurred for both T_M, which dropped by at least 40°C, and ΔH_F, which decreased by 30 – 50 %. Hairpin stability was strongly dependent on the length of the CHR, putting CHR disengagement as a step along the unfolding pathway of the hairpins. Residual helicity after the melt in DPC correlated with the proportion of NHR in the total sequence, suggesting retention of NHR helix, possibly by association of the amphiphilic helices with micelle surfaces. From the data, we reasoned that interaction of the reverse hairpin with lipids during fusion could draw the CHR away from the NHR, leading to fully exposed NHR as the fusion inhibitor. A construct in which the hairpin was fixed in place by a disulfide bond connecting CHR to the N-terminal half of the NHR lost orders of magnitude in potency. A 50% increase in the number of charged residues (with no change in overall charge) in second generation CHR did not alter antiviral potency, which could be explained by a model in which amphiphilic CHR was oriented parallel to the bilayer surface. Lipid bilayer association of NHR, buttressed by observation of a 4-fold loss of potency in constructs with higher negative charge in the NHR, could be a factor in stabilizing the helical structure and exposing the inhibitory epitope. This would imply that the endogenous gp41 prefusion ectodomain helices are also within or near the lipid bilayer[50, 60, 61].

To study the contribution of the HP epitope to antiviral activity, we prepared an NHR adduct in which a small molecule was covalently bound to the HP lysine residue. The adduct displayed a dramatic loss of potency. XFMS experiments measuring solvent accessibility changes when the ligand was attached were consistent with a pose in which the small molecule occluded the pocket, with close contacts to H52, W59 and the N-terminal end of the folded CHR domain. The definition of ligand orientation by this method, while not providing a high

resolution structure, adds to the limited body of experimental evidence of ligand association in the HP.

It is unclear from the available data whether the reverse hairpins act as trimers, targeting endogenous gp41 CHR, or as monomers, targeting endogenous gp41 NHR. Both mechanisms could be at play, since there are no disulfide bonds or other structural constraints forcing the hairpins into a trimeric structure. Clearly the trimeric structure is stable in solution and in crystalline form, but not as stable in the presence of DPC. The caveat is that DPC solutions are not an ideal mimic of bilayer membranes, as they are known to destabilize and denature α-helical membrane proteins [46], and here destabilized NHR trimer. The fact that 26-, 24- and 22-residue CHR constructs all had similar antiviral activity, yet differed in their trimer melting temperature in DPC suggests that DPC is not accurately reflecting the in situ condition. The first two had $T_M >> 37°C$ and the last had $T_M < 37°C$ (Table 3). Since the HP is formed by trimerization, the loss of potency displayed by the HP-occluding NHR-adducts would also imply that the trimer was the more likely determinant of antiviral activity. It is however conceivable that monomeric NHR-adduct could associate with endogenous gp41 NHR and disrupt further association with endogenous CHR. In either case, the presence of a folded CHR domain, even though truncated, helped with trimer stability, ameliorated aggregation of NHR surfaces and allowed the inhibitor to reach its intended target.

Antiviral and cell-cell fusion data displayed the same trends, although the effects of the disulfide bond clamp and the HP-occluding adduct were amplified in the antiviral assay compared to the cell-cell fusion assay. This could point to differences in membrane-associated prefusion gp41 or inhibitor structures or simply be a by-product of variable Env expression in the cell-cell fusion assay. Clearly, interactions along the full length of the NHR, including both N

and C-terminal sections NHR-N and NHR-C [29] played a critical role in anti-fusion activity. When the hairpin was fixed in place by a disulfide bond connecting CHR to NHR-N, or when the HP of NHR-C was occluded by a small molecule binder tethered to the pocket lysine, orders of magnitude reduction in potency were observed. This was at odds with recent data showing that a construct containing only the C-terminal section (including the HP) was inactive in an antiviral assay [29]. Possible alterations of solution behavior of the isolated C-terminal half of the NHR or its ability to reach its in situ viral target in the absence of the rest of the NHR may have caused this observed effect. Also possible is a different mechanism of action in which the reverse hairpin monomer is the active component. Either way, HP interactions were confirmed to be a critical part of the inhibitory activity of the hairpins. Further studies of the reverse hairpins, including engineered disulfide bonds to stabilize the trimer[17], may shed more light on their mechanism of action and yield enhanced antiviral activity.

MATERIALS AND METHODS

Protein Expression and Purification To maximize expression level, proteins were grown in E. coli as longer constructs containing a domain that covered the hydrophobic pocket and contained a His tag for Ni column purification [33]. The gene for protein pre-C26'(L4)N50 (Figure 2) was cloned into pET21a plasmid between EcoR1 and Nde1 restriction sites. Following synthesis in E. coli BL21(DE3) competent cells and purification using nickel affinity chromatography, the C-terminal residues covering the HP were removed by chemical cleavage at a designed Asp-Pro sequence in the constructs using 5% formic acid in 6 M guanidinium chloride at 55°C for 35 – 48 hours. An additional Ni column purification separated the His

tagged fragment and any uncleaved precursor from the final product, which was exchanged into 10 mM formate, pH 3, at which it remained stable indefinitely as a concentrated stock (1mM). Purity and molecular weight were assessed using SDS protein gel electrophoresis and Maldi – Mass Spec. Maldi data were obtained on an AB Sciex TOF/TOF Series Explorer 7000. A small amount of impurity occurred in some samples, where chemical cleavage occurred at an Asp-Ser linkage, resulting in a protein containing only the NHR; this typically represented less than 5% of the total sample.

Protein Crystallization Crystal conditions were established using an Art Robbins Instruments Crystal Phoenix Robot, screening several Hampton screens, including Index, Berkeley, PEGRX, MCSG-1 and PegIon. Crystals of C28(L4)N50 ≥ 10 μm diameter grew in several conditions and were optimized in 200mM ammonium sulfate, 100mM sodium acetate, 25% w/v PEG 4000 at pH 4.6 to 80 μm diamond-shaped crystals. Crystals of C26'(L4)N50 and C24'(L4)N50 were optimized in 0.1M lithium sulfate, 0.1M sodium citrate trihydrate at pH 6.4 and 25% PEG-1500. Prior to freezing, all crystals were dipped briefly in precipitant buffer containing 20% glycerol. Several small molecule ligands known to bind to the HP were tested in co-crystallization trials and in soaking experiments, using several indole-containing small molecules as equilibrium binders or modified to act as covalent inhibitors (Supplementary Data S2). There was insufficient electron density to describe a ligand bound in the pocket in all of these studies. Protein adducts with covalent ligands did not crystallize in any of the conditions tested.

X-ray data collection and structure determination

X-ray data was collected at the Berkeley Center for Structural Biology on beamlines 5.0.1 and 5.0.2 of the Advanced Light Source at Lawrence Berkeley National Laboratory. Diffraction data were recorded using an Dectris Pilatus 6M detector (Dectris USA Inc., Philadelphia, PA). Processing of image data was performed using the Xia2 [62]. Phases were calculated by molecular-replacement with the program PHASER [63], using the structure of GP41 mimetic (PDB id: 2XRA) [64] as a search model. Manual building using Coot [65] was alternated with reciprocal space refinement using *phenix.refine* [66]. Water molecules were automatically placed using PHENIX [67] and manually added or deleted with Coot according to peak height (3.0σ in the *Fo-Fc* map) and distance from a potential hydrogen bonding partner (< 3.5 Å). TLS refinement [68] using 2 groups per molecule, chosen using the TLSMD web server [69], was used in later rounds of refinement. All data collection, phasing, and refinement statistics are summarized in Table S1 (Supplementary Data).

Determination of free thiols DTNB reagent was freshly prepared using 2mM DTNB, 1M Tris, pH 8.0 and water in the volumetric ratio 3 : 6 : 50. A typical sample was prepared by mixing 118 μM DTNB reagent and 2 μM protein. Absorbance at 412 nm was measured after a five minute incubation time. Serial dilution of cysteine stock solutions starting at 10 mM was used to calibrate the assay. Thiol concentrations in unknowns was measured either using the cysteine calibration line or using the extinction coefficient $\varepsilon_{412} = 13.6\ mM^{-1}\ cm^{-1}$. The assay was sensitive to stock protein concentrations in the range 0.2 – 5 mM.

Protein – small molecule adduct formation Compound **2** (Supplementary Data S6) was used to form adducts with the pocket lysine (Lys-574 in HXB2 numbering) of C26'(L4)N50 and C22'(L4)N50 by reaction of 100 μM protein with an activated ester form of **2**, **2**-STP at a

stoichiometry of 1:1 and 1:3, respectively overnight at 37°C in 50 mM phosphate buffer at pH 8. The resulting adducts were concentrated and exchanged into 10 mM formate buffer using a spin filter (Amicon Ultra – 0.5ml, 3K cutoff, Sigma). The concentration of the protein – small molecule adduct was calculated using the sum of extinction coefficients of protein and small molecule at 280 nm, $e_{280} = 6.97 + 21.42\ mM^{-1}$. The formation of the adduct was checked by Maldi-MS to ensure complete coupling had occurred.

CD measurements CD spectra were obtained on a Jasco J-815 spectropolarimeter equipped with a cell positioner, peltier, water bath and titrator. Proteins with the new CHR sequence were measured at 20 μM concentration in PBS buffer, with or without 10 mM DPC. CD spectra were recorded with a 1 nm bandwidth and 0.5 nm resolution from 260 – 190 nm at room temperature. After subtracting the spectrum of buffer, the percent helicity was determined from the mean residue ellipticity (MRE), assuming 100% helix at MRE = -33000 $deg.cm^2.dmol^{-1}$ at 222nm. Thermal denaturation data were obtained by measuring θ_{222} at 0.5°C increments between 22 – 90°C with a 30s equilibration time at each step. The data from CD melts were fit assuming a two state transition from folded trimer to unfolded monomer, as published [40, 41]. The calculated data were corrected for observed pre- or post-transition linear changes in the ellipticity as a function of temperature.

The folded to unfolded transition $F_3 <—> 3U$, for protein at concentration P_t has an equilibrium folding constant

$$K = [F]/[U]^3 = \exp(-\Delta G_F/RT) – \ln (0.75P_t^2), \quad (1)$$

where ΔG_F is the folding free energy:

$$\Delta G_F = \Delta H_F * (1-T/T_M) - \Delta Cp *((T_M-T)+T\ln(T/T_M)\ ). \quad (2)$$

T_M is the temperature at the midpoint of the transition, ΔH_F is the enthalpy of folding and ΔCp is the change in heat capacity between folded and unfolded states. The observed ellipticity as a function of temperature was fit to the equation

$$\theta_{obs} = \theta_U + \alpha * (\theta_F - \theta_U). \quad (3)$$

α is the fraction folded, determined from K and P_t [41]. The parameters ΔH_F, T_M, and the ellipticity of folded and unfolded forms, θ_F and θ_U, were obtained from the fit. ΔCp values < 1 kcal/mol.K were obtained and contributed little to the overall data fitting.

XFMS materials and methods

Protein samples at 10 μM concentration in 10 mM phosphate buffer were irradiated at the Advanced Light Source (ALS) beamline 3.3.1, a 1.3 Tesla bending magnet beamline with critical energy of 3100eV and focused beam size of ~80 μm. Prior to irradiation of protein samples, beam flux was characterized using 5 μM Alexa 488 fluorescence dye (Thermo Fisher Scientific) in 10 mM phosphate buffer as previously described [70]. Protein samples pre-loaded into syringes were placed in the syringe pump capillary flow X-ray footprinting instrument [54]. Irradiation time was varied between 250 and 1500 μsec as determined by flow speed [71], and exposed samples were collected in tubes containing methionine amide to immediately quench any secondary radical reactions. Samples were stored at -80°C. Using a 1:20 (w/w) protease-to-protein ratio, the footprinted samples were digested with trypsin/Lys-C (Promega) and chymotrypsin (Promega) to ensure complete sequence coverage. Digested samples were desalted using C18 spin columns (Pierce) and peptides were resuspended in 0.1% formic acid for LC-MS/MS analysis.

LC-MS/MS analysis of peptides was conducted on an Orbitrap Exploris 480 mass spectrometer (Thermo Fisher Scientific) coupled to an Agilent 1290 UHPLC system (Agilent Technologies, Santa Clara, CA). Peptides were separated on a InfinityLab Poroshell 120 EC-C18 column (2.1 x 100 mm, 1.9 μm particle size, operated at 60°C) at a 0.400 mL/min flow rate and eluted with the following gradient: initial condition was 98% solvent A (0.1% formic acid) and 2% solvent B (99.9% acetonitrile, 0.1% formic acid). Solvent B was increased to 10% over 1.5 min, and then increased to 35% over 10 min, then up to 80% over 0.5 min, and held for 1.5 min at a flow rate of 0.6 mL/min, followed by a ramp back down to 2% B over 0.5 min where it was held for re-equilibrating the column to original conditions. The mass spectrometer was operated with the following settings: Full scan Orbitrap resolution at 60,000; AGC Target at 3.0e6; maximum injection time after 60 ms; Top 10 intense ions were isolated for HCD fragmentation per MS scan with collision energy set to 30% and intensity threshold at 5.0e3; dynamic exclusion duration set at 2 s; data-dependent MS2 scan Orbitrap resolution at 15,000; AGC target at 1.0e5; and maximum injection time after 50 ms. Data analysis was performed using PMI Byos® v.5.3.44 (Protein Metrics Inc.). Modifications searched for were the most commonly observed variable +14, +16, +32, +48, and -30 Da oxidations. Manual validation of retention-time-specific MS/MS spectra showing a high degree of fragment-ion coverage enabled confident assignment of residue-specific modifications. The extracted-ion-chromatogram peak areas of the modified and native peptides formed the basis for the quantification of modifications. The fraction unmodified for each peptide was calculated as the ratio of the integrated peak area of the unmodified peptide to the sum of integrated peak areas from the modified and unmodified peptides, and the fraction unmodified was normalized against any background oxidation seen in the unexposed control sample. The fraction of unmodified protein as a function of exposure time

was plotted in Origin v.2019b (OriginLab) and the dose-response profiles were fitted to the single exponential function $y = e^{-kt}$. The rate constant, $k(s^{-1})$, is a measure of the intrinsic hydroxyl radical reactivity and the solvent accessibility of the residue, while the ratio (R) of rate constants provides a measure of the relative change in the solvent accessibility of the residue between the free and ligand-bound states of the protein.

Antiviral and cell-cell fusion assays Cell-cell fusion and virus-cell fusion experiments were conducted as previously described [37]. Cell lines and plasmids were obtained through the NIH HIV Reagent program, Division of AIDS, NIAID, NIH, now managed by the American Type Culture Collection. Env expression vectors were donated by C. Weiss, FDA.

Briefly, target cells used were TZM-bl cells expressing CD4, CXCR4 and CCR5 [72] and containing an integrated reporter gene for firefly luciferase under control of HIV-1 LTR [73]. For cell-cell fusion experiments, effector cells were HL2/3 which produce HXB2 Env, Tat and Rev [74]. 2 x 10^4 TZM-bl cells per well of a 96 well plate were grown overnight in DMEM, then the medium replaced with serum-reduced medium containing 8 x 10^4 HL2/3 cells per well in the presence or absence of serially diluted protein inhibitors, except for negative controls in which no HL2/3 cells were added. After 5 hours, cells were lysed and luciferase expression was measured on a Biotek Synergy 2 plate reader using Luciferase Assay Reagent (Promega) according to the manufacturer's instructions.

Virus-cell fusion experiments were conducted using pseudotyped virus to measure infectivity. Pseudotyped virus was prepared in 293T cells using Env deficient HIV-1 backbone vector pSG3Δenv [75] and Env expression vector pSM-HXB2-WTgp160 using FuGENE HD transfection

reagent (Promega). Virus at 10 * $TCID_{50}$ concentration was added to TZM-Bl cells in the presence of serially diluted protein inhibitors for 5 hours in serum-reduced medium, after which the medium was replaced with DMEM complete medium. Cells were lysed and luciferase expression read after a further 24 hours. Positive and negative controls included wells with buffer only and wells without virus, respectively. Additionally, selectivity for fusion was distinguished from potential toxicity of the inhibitors using a resazurin assay (Presto Blue). The control for 100% toxicity was achieved using 15% DMSO in one or two wells of the plate.

Accession numbers

PDB: 9ARP, 9ARN, 8W37, 8W32, 8W2Y

Acknowledgments

This work was supported in part by NIH R21AI140904, NIH 1R01GM126218 and NIH P30 GM124169. Additional support was provided by Touro University California College of Osteopathic Medicine. Work conducted at the Advanced Light Source and the Molecular Foundry (Office of Basic Energy Sciences) and at the Joint BioEnergy Institute (Office of Biological and Environmental Research), was supported by the Office of Science of the U.S. DOE under contract DE-AC02-05CH11231. K. H. L. was supported by the Dr. Miriam and Sheldon G. Adelson Medical Research Foundation.

We thank Dr. Al Burlingame for access to the University of California San Francisco Mass Spectrometry Facility for protein adduct characterization. Molecular graphics and analyses were

performed with ChimeraX, developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from National Institutes of Health R01-GM129325 and the Office of Cyber Infrastructure and Computational Biology, National Institute of Allergy and Infectious Diseases, or using Chimera, developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from NIH P41-GM103311.

References

1. Chan, D. C.; Fass, D.; Berger, J. M.; Kim, P. S., Core structure of gp41 from the HIV envelope glycoprotein. *Cell* **1997,** *89* (2), 263-273.
2. Tan, K.; Liu, J.; Wang, J.; Shen, S.; Lu, M., Atomic structure of a thermostable subdomain of HIV-1 gp41. *Proceedings of the National Academy of Sciences of the United States of America* **1997,** *94* (23), 12303-8.
3. Weissenhorn, W.; Dessen, A.; Harrison, S. C.; Skehel, J. J.; Wiley, D. C., Atomic structure of the ectodomain from HIV-1 gp41 [see comments]. *Nature* **1997,** *387* (6631), 426-30.
4. Melikyan, G. B.; Markosyan, R. M.; Hemmati, H.; Delmedico, M. K.; Lambert, D. M.; Cohen, F. S., Evidence that the transition of HIV-1 gp41 into a six-helix bundle, not the bundle configuration, induces membrane fusion. [Comment In: J Cell Biol. 2000 Oct 16;151(2):F9-14]. *J Cell Biol* **2000,** *151* (2), 413-23.
5. Root, M. J.; Steger, H. K., HIV-1 gp41 as a target for viral entry inhibition. *Curr Pharm Des* **2004,** *10* (15), 1805-25.
6. Gochin, M.; Zhou, G., Amphipathic properties of HIV-1 gp41 fusion inhibitors. *Curr Topics Med Chem* **2011,** *11* (24), 3022-32.
7. Naider, F.; Anglister, J., Peptides in the treatment of AIDS. *Curr Opin Struct Biol* **2009,** *19* (4), 473-82.
8. Maeda, K.; Das, D.; Kobayakawa, T.; Tamamura, H.; Takeuchi, H., Discovery and Development of Anti-HIV Therapeutic Agents: Progress Towards Improved HIV Medication. *Curr Top Med Chem* **2019,** *19* (18), 1621-1649.
9. Eckert, D. M.; Kim, P. S., Design of potent inhibitors of HIV-1 entry from the gp41 N-peptide region. *Proc Natl Acad Sci U S A* **2001,** *98* (20), 11187-92.
10. Bewley, C. A.; Louis, J. M.; Ghirlando, R.; Clore, G. M., Design of a Novel Peptide Inhibitor of HIV Fusion That Disrupts the Internal Trimeric Coiled-coil of gp41. *J. Biol. Chem.* **2002,** *277* (16), 14238-45.
11. Zhuang, M.; Wang, W.; De Feo, C. J.; Vassell, R.; Weiss, C. D., Trimeric, coiled-coil extension on peptide fusion inhibitor of HIV-1 influences selection of resistance pathways. *Journal of Biological Chemistry* **2012,** *287* (11), 8297-309.
12. Root, M. J.; Kay, M. S.; Kim, P. S., Protein design of an HIV-1 entry inhibitor. *Science* **2001,** *291* (5505), 884-8.
13. Crespillo, S.; Camara-Artigas, A.; Casares, S.; Morel, B.; Cobos, E. S.; Mateo, P. L.; Mouz, N.; Martin, C. E.; Roger, M. G.; El Habib, R.; Su, B.; Moog, C.; Conejero-Lara, F., Single-chain protein mimetics of the N-terminal heptad-repeat region of gp41 with potential as anti-HIV-1 drugs. *Proc Natl Acad Sci U S A* **2014,** *111* (51), 18207-12.
14. Weissenhorn, W.; Calder, L. J.; Dessen, A.; Laue, T.; Skehel, J. J.; Wiley, D. C., Assembly of a rod-shaped chimera of a trimeric GCN4 zipper

and the HIV-1 gp41 ectodomain expressed in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America* **1997,** *94* (12), 6065-9.
15. Louis, J. M.; Nesheiwat, I.; Chang, L.; Clore, G. M.; Bewley, C. A., Covalent Trimers of the Internal N-terminal Trimeric Coiled-coil of gp41 and Antibodies Directed against Them are Potent Inhibitors of HIV Envelope-mediated Cell Fusion. *J Biol Chem* **2003,** *278*, 20278-85.
16. Louis, J. M.; Bewley, C. A.; Clore, G. M., Design and Properties of N_{CCG}-gp41, a Chimeric gp41 Molecule with Nanomoar HIV Fusion Inhibitory Activity. *J. BIological Chemistry* **2001,** *276* (31), 29485-89.
17. Bianchi, E.; Finotto, M.; Ingallinella, P.; Hrin, R.; Carella, A. V.; Hou, X. S.; Schleif, W. A.; Miller, M. D.; Geleziunas, R.; Pessi, A., Covalent stabilization of coiled coils of the HIV gp41 N region yields extremely potent and broad inhibitors of viral infection. *Proc Natl Acad Sci U S A* **2005,** *102* (36), 12903-8.
18. Dwyer, J. J.; Wilson, K. L.; Martin, K.; Seedorff, J. E.; Hasan, A.; Medinas, R. J.; Davison, D. K.; Feese, M. D.; Richter, H. T.; Kim, H.; Matthews, T. J.; Delmedico, M. K., Design of an engineered N-terminal HIV-1 gp41 trimer with enhanced stability and potency. *Protein Science* **2008,** *17* (4), 633-43.
19. Greenberg, M. L.; Cammack, N., Resistance to enfuvirtide, the first HIV fusion inhibitor. *J Antimicrob Chemother* **2004,** *54* (2), 333-40.
20. Reeves, J. D.; Lee, F. H.; Miamidian, J. L.; Jabara, C. B.; Juntilla, M. M.; Doms, R. W., Enfuvirtide resistance mutations: impact on human immunodeficiency virus envelope function, entry inhibitor sensitivity, and virus neutralization. *J Virol* **2005,** *79* (8), 4991-9.
21. Rimsky, L. T.; Shugars, D. C.; Matthews, T. J., Determinants of human immunodeficiency virus type 1 resistance to gp41-derived inhibitory peptides. *Journal of Virology* **1998,** *72* (2), 986-93.
22. Desmezieres, E.; Gupta, N.; Vassell, R.; He, Y.; Peden, K.; Sirota, L.; Yang, Z.; Wingfield, P.; Weiss, C. D., Human immunodeficiency virus (HIV) gp41 escape mutants: cross-resistance to peptide inhibitors of HIV fusion and altered receptor activation of gp120. *J Virol* **2005,** *79* (8), 4774-81.
23. Wang, W.; De Feo, C. J.; Zhuang, M.; Vassell, R.; Weiss, C. D., Selection with a peptide fusion inhibitor corresponding to the first heptad repeat of HIV-1 gp41 identifies two genetic pathways conferring cross-resistance to peptide fusion inhibitors corresponding to the first and second heptad repeats (HR1 and HR2) of gp41. *Journal of Virology* **2011,** *85* (24), 12929-38.
24. Kahle, K. M.; Steger, H. K.; Root, M. J., Asymmetric deactivation of HIV-1 gp41 following fusion inhibitor binding. *PLoS Pathog* **2009,** *5* (11), e1000674.
25. Steger, H. K.; Root, M. J., Kinetic dependence to HIV-1 entry inhibition. *J Biol Chem* **2006,** *281* (35), 25813-21.
26. Jurado, S.; Cano-Munoz, M.; Morel, B.; Standoli, S.; Santarossa, E.; Moog, C.; Schmidt, S.; Laumond, G.; Camara-Artigas, A.; Conejero-Lara, F.,

Structural and Thermodynamic Analysis of HIV-1 Fusion Inhibition Using Small gp41 Mimetic Proteins. *J Mol Biol* **2019,** *431* (17), 3091-3106.
27. Lee, J. H.; Ozorowski, G.; Ward, A. B., Cryo-EM structure of a native, fully glycosylated, cleaved HIV-1 envelope trimer. *Science* **2016,** *351* (6277), 1043-8.
28. Jurado, S.; Cano-Munoz, M.; Polo-Megias, D.; Conejero-Lara, F.; Morel, B., Thermodynamic dissection of the interface between HIV-1 gp41 heptad repeats reveals cooperative interactions and allosteric effects. *Arch Biochem Biophys* **2020,** *688*, 108401.
29. Jurado, S.; Moog, C.; Cano-Munoz, M.; Schmidt, S.; Laumond, G.; Ruocco, V.; Standoli, S.; Polo-Megias, D.; Conejero-Lara, F.; Morel, B., Probing Vulnerability of the gp41 C-Terminal Heptad Repeat as Target for Miniprotein HIV Inhibitors. *J Mol Biol* **2020,** *432* (20), 5577-5592.
30. Eckert, D. M.; Malashkevich, V. N.; Hong, L. H.; Carr, P. A.; Kim, P. S., Inhibiting HIV-1 entry: discovery of D-peptide inhibitors that target the gp41 coiled-coil pocket. *Cell* **1999,** *99* (1), 103-15.
31. Welch, B. D.; Francis, J. N.; Redman, J. S.; Paul, S.; Weinstock, M. T.; Reeves, J. D.; Lie, Y. S.; Whitby, F. G.; Eckert, D. M.; Hill, C. P.; Root, M. J.; Kay, M. S., Design of a potent D-peptide HIV-1 entry inhibitor with a strong barrier to resistance. *J Virol* **2010,** *84* (21), 11235-44.
32. Chong, H.; Qiu, Z.; Su, Y.; Yang, L.; He, Y., Design of a highly potent HIV-1 fusion inhibitor targeting the gp41 pocket. *AIDS* **2015,** *29* (1), 13-21.
33. Walsh, J. D.; Chu, S.; Zhang, S. Q.; Gochin, M., Design and characterization of swapped-domain constructs of HIV-1 glycoprotein-41 as receptors for drug discovery. *Protein Eng Des Sel* **2015,** *28* (4), 107-16.
34. Chu, S.; Kaur, H.; Nemati, A.; Walsh, J. D.; Partida, V.; Zhang, S. Q.; Gochin, M., Swapped-Domain Constructs of the Glycoprotein-41 Ectodomain Are Potent Inhibitors of HIV Infection. *ACS Chem Biol* **2015,** *10* (5), 1247-57.
35. Otaka, A.; Nakamura, M.; Daisuke, N.; Kodama, E.; Uchiyama, S.; Nakamura, H.; Kobayashi, Y.; Matsuoka, M.; Fujii, N., Remodelling of gp41-C34 Peptide Leads to Highly Effective Inhibitors of the Fusion of HIV-1 with Target Cells. *Angew. Chem. Int. Ed.* **2002,** *41* (16), 2938-9.
36. He, L.; Zhou, G.; Sofiyev, V.; Garcia, E.; Nguyen, N.; Li, K. H.; Gochin, M., Targeting a Conserved Lysine in the Hydrophobic Pocket of HIV-1 gp41 Improves Small Molecule Antiviral Activity. *Viruses* **2022,** *14* (12), 2703.
37. Zhou, G.; He, L.; Li, K. H.; Pedroso, C. C. S.; Gochin, M., A targeted covalent small molecule inhibitor of HIV-1 fusion. *Chem Commun (Camb)* **2021,** *57* (37), 4528-4531.
38. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E., UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **2004,** *25* (13), 1605-12.
39. Crooks, R. O.; Rao, T.; Mason, J. M., Truncation, randomization, and selection: generation of a reduced length c-Jun antagonist that retains high interaction stability. *J Biol Chem* **2011,** *286* (34), 29470-9.

40. Greenfield, N. J., Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions. *Nat Protoc* **2006,** *1* (6), 2527-35.
41. Greenfield, N. J., Circular dichroism analysis for protein-protein interactions. *Methods Mol Biol* **2004,** *261*, 55-78.
42. Robertson, A. D.; Murphy, K. P., Protein Structure and the Energetics of Protein Stability. *Chemical Reviews* **1997,** *97* (5), 1251-1268.
43. Miles, A. J.; Wallace, B. A., Circular dichroism spectroscopy of membrane proteins. *Chem Soc Rev* **2016,** *45* (18), 4859-72.
44. le Maire, M.; Champeil, P.; Moller, J. V., Interaction of membrane proteins and lipids with solubilizing detergents. *Biochimica et Biophysica Acta* **2000,** *1508* (1-2), 86-111.
45. Zhou, G.; Chu, S.; Kohli, A.; Szoka, F. C.; Gochin, M., Biophysical studies of HIV-1 glycoprotein-41 interactions with peptides and small molecules - Effect of lipids and detergents. *Biochim Biophys Acta Gen Subj* **2020,** *1864* (12), 129724.
46. Chipot, C.; Dehez, F.; Schnell, J. R.; Zitzmann, N.; Pebay-Peyroula, E.; Catoire, L. J.; Miroux, B.; Kunji, E. R. S.; Veglia, G.; Cross, T. A.; Schanda, P., Perturbations of Native Membrane Protein Structure in Alkyl Phosphocholine Detergents: A Critical Assessment of NMR and Biophysical Studies. *Chemical Reviews* **2018,** *118* (7), 3559-3607.
47. Klug, Y. A.; Rotem, E.; Schwarzer, R.; Shai, Y., Mapping out the intricate relationship of the HIV envelope protein and the membrane environment. *Biochim Biophys Acta Biomembr* **2017,** *1859* (4), 550-560.
48. Aisenbrey, C.; Bechinger, B., Structure, interactions and membrane topology of HIV gp41 ectodomain sequences. *Biochim Biophys Acta Biomembr* **2020,** *1862* (7), 183274.
49. Lakomek, N. A.; Kaufman, J. D.; Stahl, S. J.; Louis, J. M.; Grishaev, A.; Wingfield, P. T.; Bax, A., Internal dynamics of the homotrimeric HIV-1 viral coat protein gp41 on multiple time scales. *Angewandte Chemie. International Ed. In English* **2013,** *52* (14), 3911-5.
50. Roche, J.; Louis, J. M.; Grishaev, A.; Ying, J.; Bax, A., Dissociation of the trimeric gp41 ectodomain at the lipid-water interface suggests an active role in HIV-1 Env-mediated membrane fusion. *Proceedings of the National Academy of Sciences of the United States of America* **2014,** *111* (9), 3425-30.
51. Craig, D. B.; Dombkowski, A. A., Disulfide by Design 2.0: a web-based tool for disulfide engineering in proteins. *BMC Bioinformatics* **2013,** *14*, 346.
52. Krittanai, C.; Johnson, W. C., Correcting the circular dichroism spectra of peptides for contributions of absorbing side chains. *Anal Biochem* **1997,** *253* (1), 57-64.
53. Xu, G.; Chance, M. R., Hydroxyl radical-mediated modification of proteins as probes for structural proteomics. *Chem Rev* **2007,** *107* (8), 3514-3543.

54. Gupta, S.; Celestre, R.; Petzold, C. J.; Chance, M. R.; Ralston, C., Development of a microsecond X-ray protein footprinting facility at the Advanced Light Source. *J Synchrotron Radiat* **2014,** *21* (Pt 4), 690-699.
55. Kiselar, J.; Chance, M. R., High-Resolution Hydroxyl Radical Protein Footprinting: Biophysics Tool for Drug Discovery. *Annu Rev Biophys* **2018,** *47*, 315-333.
56. Zhou, G.; Chu, S.; Nemati, A.; Huang, C.; Snyder, B. A.; Ptak, R. G.; Gochin, M., Investigation of the molecular characteristics of bisindole inhibitors as HIV-1 glycoprotein-41 fusion inhibitors. *Eur J Med Chem* **2019,** *161*, 533-542.
57. Hacker, S. M.; Backus, K. M.; Lazear, M. R.; Forli, S.; Correia, B. E.; Cravatt, B. F., Global profiling of lysine reactivity and ligandability in the human proteome. *Nat Chem* **2017,** *9* (12), 1181-1190.
58. Bianco, G.; Forli, S.; Goodsell, D. S.; Olson, A. J., Covalent docking using autodock: Two-point attractor and flexible side chain methods. *Protein Sci* **2016,** *25* (1), 295-301.
59. Eberhardt, J.; Santos-Martins, D.; Tillack, A. F.; Forli, S., AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. *J Chem Inf Model* **2021,** *61* (8), 3891-3898.
60. Lakomek, N. A.; Kaufman, J. D.; Stahl, S. J.; Wingfield, P. T., HIV-1 envelope protein gp41: an NMR study of dodecyl phosphocholine embedded gp41 reveals a dynamic prefusion intermediate conformation. *Structure* **2014,** *22* (9), 1311-21.
61. Roche, J.; Louis, J. M.; Aniana, A.; Ghirlando, R.; Bax, A., Complete dissociation of the HIV-1 gp41 ectodomain and membrane proximal regions upon phospholipid binding. *Journal of Biomolecular NMR* **2015,** *61* (3-4), 235-48.
62. Otwinowski, Z.; Minor, W., Processing of X-ray diffraction data collected in oscillation mode. *Method Enzymol* **1997,** *276*, 307-326.
63. McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J., Phaser crystallographic software. *J Appl Crystallogr* **2007,** *40* (Pt 4), 658-674.
64. She, J.; Han, Z.; Kim, T. W.; Wang, J.; Cheng, W.; Chang, J.; Shi, S.; Yang, M.; Wang, Z. Y.; Chai, J., Structural insight into brassinosteroid perception by BRI1. *Nature* **2011,** *474* (7352), 472-6.
65. Emsley, P.; Cowtan, K., Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **2004,** *60* (Pt 12 Pt 1), 2126-32.
66. Afonine, P. V.; Grosse-Kunstleve, R. W.; Echols, N.; Headd, J. J.; Moriarty, N. W.; Mustyakimov, M.; Terwilliger, T. C.; Urzhumtsev, A.; Zwart, P. H.; Adams, P. D., Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr* **2012,** *68* (Pt 4), 352-67.
67. Liebschner, D.; Afonine, P. V.; Baker, M. L.; Bunkoczi, G.; Chen, V. B.; Croll, T. I.; Hintze, B.; Hung, L. W.; Jain, S.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R. D.; Poon, B. K.; Prisant, M. G.; Read, R. J.; Richardson, J. S.; Richardson, D. C.; Sammito, M. D.; Sobolev, O. V.; Stockwell, D. H.;

Terwilliger, T. C.; Urzhumtsev, A. G.; Videau, L. L.; Williams, C. J.; Adams, P. D., Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr D Struct Biol* **2019,** *75* (Pt 10), 861-877.
68. Painter, J.; Merritt, E. A., Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr D Biol Crystallogr* **2006,** *62* (Pt 4), 439-50.
69. Painter, J.; Merritt, E. A., TLSMD web server for the generation of multi-group TLS models. *J Appl Crystallogr* **2006,** *39*, 109-111.
70. Gupta, S.; Sullivan, M.; Toomey, J.; Kiselar, J.; Chance, M. R., The Beamline X28C of the Center for Synchrotron Biosciences: a national resource for biomolecular structure and dynamics experiments using synchrotron footprinting. *J Synchrotron Radiat* **2007,** *14* (Pt 3), 233-43.
71. Gupta, S.; Chen, Y.; Petzold, C. J.; DePonte, D. P.; Ralston, C. Y., Development of Container Free Sample Exposure for Synchrotron X-ray Footprinting. *Anal Chem* **2020,** *92* (1), 1565-1573.
72. Platt, E. J.; Wehrly, K.; Kuhmann, S. E.; Chesebro, B.; Kabat, D., Effects of CCR5 and CD4 cell surface concentrations on infections by macrophagetropic isolates of human immunodeficiency virus type 1. *J Virol* **1998,** *72* (4), 2855-64.
73. Wei, X.; Decker, J. M.; Liu, H.; Zhang, Z.; Arani, R. B.; Kilby, J. M.; Saag, M. S.; Wu, X.; Shaw, G. M.; Kappes, J. C., Emergence of resistant human immunodeficiency virus type 1 in patients receiving fusion inhibitor (T-20) monotherapy. *Antimicrob Agents Chemother* **2002,** *46* (6), 1896-905.
74. Ciminale, V.; Felber, B. K.; Campbell, M.; Pavlakis, G. N., A bioassay for HIV-1 based on Env-CD4 interaction. *AIDS Res Hum Retroviruses* **1990,** *6* (11), 1281-7.
75. Li, M.; Gao, F.; Mascola, J. R.; Stamatatos, L.; Polonis, V. R.; Koutsoukos, M.; Voss, G.; Goepfert, P.; Gilbert, P.; Greene, K. M.; Bilska, M.; Kothe, D. L.; Salazar-Gonzalez, J. F.; Wei, X.; Decker, J. M.; Hahn, B. H.; Montefiori, D. C., Human immunodeficiency virus type 1 env clones from acute and early subtype B infections for standardized assessments of vaccine-elicited neutralizing antibodies. *J Virol* **2005,** *79* (16), 10108-25.