Published in final edited form as:

Org Biomol Chem. 2017 June 28; 15(24): 5210–5219. doi:10.1039/c7ob00954b.

Evaluation of Ligand-Based NMR Screening Methods to Characterize Small Molecule Binding to HIV-1 glycoprotein-41

Shidong Chu^a, Guangyan Zhou^a, and Miriam Gochin^{a,b}

^aDepartment of Basic Sciences, Touro University-California, Vallejo, CA 94592

^bDepartment of Pharmaceutical Chemistry, University of California San Francisco, CA 94143

Abstract

Small molecule inhibitors of glycoprotein-41 (gp41) are able to prevent HIV infection by binding to a hydrophobic pocket (HP) contained within the gp41 ectodomain, and preventing progression of fusion. There is little structural information on gp41 – ligand complexes, owing to hydrophobicity of the ligands, occlusion of the HP in folded gp41 ectodomain, and failure to form crystals of complexes. Here we used an engineered gp41 ectodomain protein containing an exposed HP and a small molecule designed to bind with weak affinity to the HP. We evaluated NMR methods, including WaterLOGSY, Saturation Transfer Difference spectroscopy (STD-NMR) and ¹H relaxation rate difference spectroscopy with and without target irradiation (DIRECTION) for their ability to probe complex formation and structure. WaterLOGSY was the most sensitive technique for monitoring formation of the complex. STD-NMR and DIRECTION experiments gave similar pharmacophore mapping profiles, although the low dynamic range of the DIRECTION experiment limited its discrimination and sensitivity. A unique binding pose was identified from the STD data and provided clues for future optimization. Advantages and disadvantages of the techniques are discussed. This is the first example of the use of STD for structural analysis of a gp41- small molecule complex.

Graphical abstract

STD NMR pinpoints the binding mode of a small ligand in the hydrophobic pocket of gp41

Introduction

HIV-1 fusion follows viral attachment and is mediated by the membrane spanning envelope protein gp41, which undergoes a conformational rearrangement of its ectodomain in the process.^{1–5} Inhibition of fusion has been accomplished with peptides and small molecules that compete with the intrinsic refolding process.^{6–10} Small molecule inhibitors bind to a hotspot within the protein-protein interface called the hydrophobic pocket (HP),^{11–14} with binding affinity that is correlated to efficacy.^{15–17} Inhibition of HIV-1 fusion by small molecules has proved challenging in various aspects, not least of which is the lack of structural data to support ligand optimization. Like many inhibitors targeting protein – protein interactions,¹⁸ gp41 inhibitors are quite hydrophobic, which hinders biophysical studies,¹⁹ and crystal structures of gp41 – ligand complexes have not been forthcoming. Chemical elaboration has often relied on computational docking to predict the binding pose. We have obtained some experimental structural information on weakly binding ligands in fast exchange, utilizing paramagnetic relaxation NMR through a nearby spin label to provide distance constraints.^{20, 21}

In this study, we examined ligand-detected NMR techniques that do not require specific labeling of protein or ligand for their ability to identify the binding pose of a ligand **4** in the HP. We explored sample conditions with the aid of WaterLOGSY (water-ligand observed via gradient spectroscopy) NMR, a sensitive technique for probing ligand aggregation and the formation and stability of complexes with macromolecules.^{22, 23} Saturation Transfer Difference (STD) spectroscopy and ¹H relaxation rate difference spectroscopy with and without target irradiation (DIRECTION) were examined for epitope mapping and determination of binding mode.^{24–27}

STD is a well-established method for quantitative group epitope mapping (GEM) of ligands with sub- μ M to mM affinity.²⁸ Saturation transfer from protein to ligand depends on the proximity of each ligand proton to protein protons, through the intermolecular NOE. Thus variation in STD intensities suggests the degree to which each ligand position is buried in the binding site, providing information on the binding mode. Artefacts can result with measurement of STD at a single saturation time because the accumulated saturation varies with T₁ relaxation rates across the ligand. The problem can be overcome by obtaining initial rates of magnetization transfer derived from saturation buildup curves,^{28–30} or by CORCEMA-ST calculation of the buildup curves.³¹ Collection of multiple saturation time points was precluded here by the rather low stability of the complexes. Instead, we applied the GEM-CRL method of scaling STD intensities by T₁ relaxation rates.³² This simplified method did not require knowledge of the exact protein structure, or of various thermodynamic and kinetic parameters required for CORCEMA-ST calculations. The gp41 HP is known for the flexibility of surrounding side chains, making an exact CORCEMA calculation challenging.³³

DIRECTION is a relatively new technique for pharmacophore mapping, based on differences in ligand proton T₁ relaxation rates during on- or off-resonance saturation of the protein.²⁶ It is an alternative method of cancelling out relaxation rate differences, leaving in place the

intermolecular NOE term. It too relies on measurement of initial rates for accurate determination of T_1 's.

Here we report a unique binding pose of **4** in the HP obtained by STD NMR. This is the first example of the use of the STD technique applied to gp41. The DIRECTION experiment, while qualitatively similar, lacked the dynamic range to reliably evaluate the binding pose of the weakly binding ligand **4**.

Results and Discussion

Molecule Design

For the protein receptors, we utilized recently described reverse hairpin constructs of the gp41 ectodomain, in which the C-heptad repeat (CHR) preceded the N-heptad repeat (NHR) in the primary sequence, separated by a 4 residue loop.³⁴ The constructs formed a trimer of hairpins along the NHR axis. The monomers are shown in Figure 1, labeled according to the length of each domain. C39(L4)N50 contained the hydrophobic pocket binding domain (HPbd), forming an extremely stable hairpin complex unable to bind small molecules that target the pocket.³⁴ However, small molecules that bound elsewhere, such as at the C-terminal end of the NHR, could be observed to bind by WaterLOGSY NMR. In C28(L4)N50 and C21(L4)N50, the HP was exposed by truncating the CHR to remove the HPbd. CD and other (unpublished) biophysical data confirm that C28(L4)N50 and C21(L4)N50 are fully folded. Both associate with small molecule hydrophobic pocket binders, as determined by WaterLOGSY NMR.³⁴

Ligand **4** (Scheme 1, Mol Wt 329, cLogP 3.33) was designed to be soluble in aqueous buffer at neutral pH, to form a stable complex with gp41 HP and to have a good quality well-resolved 1D NMR spectrum. A N_1 -benzoic acid substituted indole formed the core scaffold, based on previous design criteria.³⁵ A 6,5'-linked pyrimidine increased size and solubility while ensuring good resonance dispersion. Peak assignment was accomplished using COSY, TOCSY and NOESY 2D spectroscopy.

4 bound exclusively in the HP of gp41

Using a fluorescence competitive inhibition assay for the hydrophobic pocket,¹⁶ we determined an inhibition constant K_I for **4** of 160 μM . One (1) mM **4** in buffer containing 2% DMSO showed no self-association or aggregation, as evidenced by the absence of signal in the WaterLOGSY NMR spectrum (Figure 2). Selective association of **4** with the HP was confirmed by WaterLOGSY experiments in the presence of receptors C39(L4)N50, C28(L4)N50 and C21(L4)N50. WaterLOGSY signal was observed only in the presence of the latter two proteins, which contain an exposed HP (Figure 2). Furthermore, **4** was displaced by a strongly binding C-peptide containing the HPbd (Supplementary Data S2). Continuous WaterLOGSY measurements showed that the complex was stable for approximately 7 hours in the NMR tube.

Four distinct docked clusters of **4** in the HP were obtained

Computational docking of **4** in PDB structure *2xra* using OpenEye FRED³⁶ docking gave four distinctly different clusters from ten saved poses (Table S3.1, Figure 3), Cluster 1 (poses 1, 2 and 5), Cluster 2 (poses 3 and 4), Cluster 3 (pose 6 and 9) and Cluster 4 (Poses 7, 8 and 10). Poses within a cluster had RMSD of heavy atoms within 0.6Å, except for cluster 3, where the RMSD between poses was 2.4Å. Poses in different clusters were > 4Å apart in RMSD (Table S3.1). There was little energetic discrimination between the calculated poses.

DIRECTION and STD experiments gave similar pharmacophore mapping profiles but DIRECTION data had lower dynamic range

STD NMR and DIRECTION experiments were conducted at 800 MHz using a **4** : C28(L4)N50 ratio of 100, with a saturation time of 3 seconds. Experimental STD intensities qSTD-AF and DIRECTION values were determined from equations 2 and 4, respectively. Figure 4 shows the STD result, with an insert showing ligand proton positions and relative STD intensities. A table of assignments, T_1 's, STD-AF and qSTD-AF values for **4** is provided in the Supplementary Information (Table S3.2). Four protons with overlapping resonances (A2, A6, C1, C3) were ignored in the subsequent analysis.

The epitope profile obtained for **4** from the STD and DIRECTION experiments is shown in Figure 5. The two patterns are qualitatively similar. The error range of the DIRECTION experiment was high relative to the small differences observed between on- and off-resonance relaxation rates. With the exception of the protons at positions B4, B7 and possibly the methylene group, the profile is relatively flat within the error range. However, the observed qualitative agreement between these two techniques is promising, and it is likely that the DIRECTION experiment will have improved dynamic range for higher affinity ligands, where there will be a larger difference between on- and off-resonance relaxation.²⁶

The methylene group was an outlier between the DIRECTION and qSTD result. Both methods are intended to cancel out the effect of the short proton relaxation time of the geminal protons, which leads to dissipation of cross-relaxation,^{32, 37} but they gave a very different outcome. Strong dipolar coupling between the geminal protons may preclude use of the single exponential model for fitting T_1 's. The GEM-CRL method for analysis of STD uses these T_1 values and may in addition underestimate the intermolecular cross-relaxation rate for the methylene group compared to STD initial rates, although the error is estimated to be no more than ~20% (Supplementary Information S4). In the DIRECTION experiment short variable delays T_{var} are used in order to approximate the single exponential model (equation 3)²⁶; however $T_{var} > T_1$ for the methylene protons, incurring spin diffusion effects and likely additional error. Because of the higher sensitivity of STD data, we focused on fitting STD profiles to evaluate the binding pose of **4** in the HP.

STD NMR data agreed with only one of the docked clusters

Qualitative analysis of the relative qSTD-AF intensities did not allow cluster discrimination by eye. Instead, we examined whether we could quantitatively discriminate between clusters by calculating relative STD factors for each pose as a sum of ligand - protein NOE's, and

comparing the results obtained with experimental STD's. We calculated $\Sigma(1/r^6)$ between each ligand proton and surrounding non-exchangeable protein protons as previously described,²⁶ where r is the interproton distance, using geometric averaging for methyl groups. STD profiles are shown in Figure 6 for several poses. It is evident that pose 1 of Cluster 1 gave an excellent match of calculated profile to that observed experimentally. A correlation coefficient of 0.73 between observed qSTD-AF values and calculated intermolecular NOE's was obtained. No correlation was observed with any of the other clusters.

Overall STD profile patterns were distinctive, but individual qSTD-AF values and NOE calculations were imprecise

qSTD-AF values for individual protons varied depending on how the data was analyzed. For example, using free ligand T_1 's instead of T_1^P values for scaling (Equation 2) resulted in up to a 2-fold change in an individual qSTD-AF value (Figure 6A). Additional approximations are associated with using the GEM-CRL scaling method.³² However, the overall profile pattern remained consistent. Concomitantly, calculated NOE values were very sensitive to the exact ligand position. Poses 1, 2 and 5 in Cluster 1 were almost overlapping (RMSD < 0.6Å, superimposed in Figure 3A), but average calculated relative intermolecular NOE's for individual protons varied between poses by up to a factor of ~2 (Figure 7A–C). The result was a fairly large standard deviation of calculated NOE's (Figure 7D). This is a reflection of the sensitivity of the calculation to atomic position. The $1/r^6$ scaling factor resulted in proton – proton distances ~ 2 Å contributing significantly to the total calculated NOE; a change of 0.2 Å in one of these short distances could alter the overall calculated NOE by 50%. The average as well as individual envelopes of calculated protein-ligand NOE's for poses 1, 2 and 5 generally showed the same pattern, with the exception of pose 2, in which the STD for A5 was smaller than the STD for B2 (Figure 7). Interestingly, in pose 1 and 5, a hydrogen bond was predicted between the carboxylate group of **4** and the pocket Arg579 ω NH₂ group, but this did not occur in pose 2 in which the ligand was shifted slightly to the right. Ultimately, compatibility was determined by observation of experimental and calculated similarity in profile pattern across the whole ligand.

STD results highlighted the possibility that degeneracy could have occurred in pose definition of **4**

Examination of the STD patterns for the Clusters 2 – 4 (Figure 6) led to several interesting observations. The lack of pattern matching with the experimental STD's was obvious, clearly delineating these clusters as deviating from the experimental data. However, we also observed that the identification of cluster 1 as the “correct” docked pose was fortuitous, since Cluster 4 and pose 6 (part of Cluster 3), which differ by > 4Å RMSD, gave similar STD patterns and would not be distinguishable experimentally given the resolution and error ranges of the method. Additionally, poses 6 and 9, which we identified as a single Cluster 3 due to a similar orientation of rings in the pocket (RMSD 2.358Å) gave significantly different STD patterns. Therefore it is not possible to predict under which exact circumstances STD can be used semi-quantitatively to reveal the binding mode for a small ligand such as **4**. The calculation was also dependent on having selected an appropriate pocket structure. The calculation was sensitive to the exact arrangement of the flexible side

chains surrounding the HP. Trying a PDB structure other than *2xra* did not produce a match of the experimental STD to any of the docked poses.

Conclusions

We have used computational docking and ligand-based NMR screening to define the binding site and pose of a weakly bound indole **4** in the hydrophobic pocket of HIV-1 gp41.

WaterLOGSY was the most sensitive method for characterizing the formation of complexes, the site of binding and the stability of the sample. A spectrum was accumulated in 15 – 30 minutes at 400 MHz. STD and DIRECTION experiments provided similar binding epitopes, from which a compatible computationally docked cluster could be selected.

STD data had higher signal to noise. Certainly the DIRECTION experiment was much faster, requiring < 15 minutes to complete at 800 MHz, while STD experiments required > 160 minutes under our conditions. The DIRECTION experiment is likely more sensitive to ligand binding affinity, requiring higher protein-ligand cross-relaxation rates for good discrimination.²⁶ Limitations to the STD experiment included variability within individual proton signals and an observed degeneracy of the epitope profile across two different bound clusters (although neither of these matched the experimental data in this case). The observed signal level for fast relaxing methylene groups was inconsistent between STD and DIRECTION methods, and interpretation remains uncertain. However, the selected docked cluster result suggested that the qSTD-AF value was a better match.

We selected PDB entry *2xra* for virtual docking because of earlier data showing it correlated well with SAR data we have accumulated on HP binding small molecules.¹⁵ This model proved successful for identifying a docked structure of **4** compatible with STD data. In this structure, the indole moiety of **4** was buried in the pocket along one edge and a hydrogen bond was predicted between the carboxylate group of the ligand and the pocket Arg579 residue (Figure 8). The indole and pyrimidine ring planes were tilted by 30°. In the final structure, we observed a potential steric clash between two hydrogen atoms on the ligand at positions A6 and C1 with protein protons on Gly 572 and Lys 574' side chain, respectively. Only a small energetic penalty was generated compared to the favorable shape, desolvation and H-bonding terms in FRED docking. This is attributed to an emphasis on ligand conformational energy and hydrogen bonding in the energy calculations. The absence of experimental data for both A6 and C1 protons resulted in no conflict in calculated vs. observed NOE's. However back-calculation at these positions yielded unreasonably large intermolecular NOE's because of short interproton distances involving A6 and C1. This suggests that *2xra*, which contains an antibody bound in the HP,³⁸ adequately but not fully represents the structure of gp41 HP with a bound small ligand.

The low potency of **4** is explained by the adopted pose, since there is minimal hydrophobic interaction within the deep pocket. Expansion of this section of the molecule is likely to yield higher affinity. Properties of the polar groups that extend outwards from the pocket can be adjusted to increase polar interactions with residues on the outer rim of the pocket. A model of the structure of **4** in the context of the whole complex is shown in the Supplementary Data (Figure S5).

We have also shown that adaptation of quantitative STD analytical methods provided a powerful approach for analyzing the binding mode in the gp41 HP, and one that could be applied even when limited stability of the ligand - protein complex restricted the amount of time available to collect data, and where there was limited information about HP structure. Data were obtained at a single saturation time with adjustment for ligand relaxation rates. For each virtual docked pose, summation of the inverse sixth power of protein – ligand interproton distances (i.e. relative intermolecular NOE) was compared to STD profile patterns to identify a unique binding mode. The analysis falls in between qualitative group epitope mapping and full relaxation matrix calculations using CORCEMA-ST, which requires detailed knowledge of the protein structure and kinetics of the exchange. Since the calculation of intermolecular NOE's was very sensitive to ligand and receptor conformation, the method depended on having an approximately correct pocket structure and validated docking software. In our case, OpenEye® FRED docking and the availability of the *2xra* PDB structure were instrumental in obtaining the final results.

Experimental

Small Molecule Synthesis

Synthesis of **4** is shown in Scheme 1. Details of the synthesis and characterization are provided in the Supplementary Information (S1).

Expression, Purification and Characterization of gp41 Proteins

Proteins C39(L4)N50, C28(L4)N50 and C21(L4)N50 were overexpressed in *E. coli* and purified using Ni affinity and size exclusion chromatography as previously reported.³⁴ Molecular weight and purity of the proteins were confirmed by LC/ESI-MS and SDS Page gel. Molecular weights of the proteins used as receptors were 33.9 kDa (C39(L4)N50), 30.0 kDa (C28(L4)N50) and 27.3 kDa (C21(L4)N50) in their trimeric form.

Competitive Inhibition Assays

Inhibition constant K_i for binding in the hydrophobic pocket was determined using a fluorescence intensity assay as previously described.^{39, 40} Briefly, trimeric NHR was formed using octahedral metal ligation to NHR peptides via covalently attached N-terminal bipyridyl groups, and ligand binding was evaluated by displacement of a fluorescent peptide probe from the HP, with a concomitant increase in fluorescence intensity.

WaterLOGSY NMR Spectroscopy

WaterLOGSY experiments were conducted on a Bruker AVANCE 400 MHz NMR spectrometer equipped with a TXI probe head. Samples were dissolved in 50 mM phosphate buffer, pH 7.4, containing 10% D₂O and 2% DMSO-d₆. Typical sample concentrations were 500 – 1000 μ M ligand and 10 μ M protein. Selective saturation of the water signal was achieved with a 20 ms Gaussian pulse. A 1.2 s NOE mixing time was used for saturation transfer, and a 3.2 ms flipback pulse and excitation sculpting with gradients⁴¹ were used for water suppression during data collection. Samples were typically recorded with 256 transients. Spectra were phased so that resonances that experience the negative NOE of the protein appear negative.

STD NMR Spectroscopy

Samples for STD NMR were similar to those used for WaterLOGSY experiments, but contained less protein, with 100 fold excess of ligand over protein, and were in 50 mM phosphate buffer in 100% D₂O. STD NMR was carried out on a Bruker AVANCE 800 MHz NMR spectrometer equipped with a TXI probe head, using a pseudo 2D pulse sequence which interleaved acquisition of on- and off-resonance saturation. Selective saturation of protein methyl groups was achieved using a train of 50 ms Gaussian pulses at 0.96 ppm. The off-resonance frequency was set to 28 ppm for the reference spectra. Residual HDO signal was suppressed by excitation sculpting with gradients. 2048 transients were obtained for each spectrum. STD difference spectra were obtained by subtracting the reference spectrum from the protein-proton saturated spectrum. The STD amplification factor⁴² was obtained with the formula

$$\text{STD-AF} = \varepsilon(I_0 - I_{\text{sat}})/I_0 \quad (1)$$

where $\varepsilon = L_t/R_t$, I_{sat} and I_0 are the peak intensities with on and off protein resonance saturation and L_t , R_t are the total concentrations of ligand and protein. STD-AF was additionally scaled by the inverse of the proton T_1 to compensate for faster dissipation of NOE effects for faster relaxing protons (GEM-CRL).³²

$$\text{qSTD-AF} = \text{STD-AF}/T_1 \quad (2)$$

We obtained T_1 for ligand alone (T_1^0) and in the presence of protein in the ratio used for STD studies (T_1^P). T_1 's of free ligands in the absence of protein were longer (See Supplementary Information Table S3.2). qSTD-AF values scaled by T_1^0 and T_1^P were averaged..

Equation 2 was validated in the original study as a method to cancel the effect of differential ligand relaxation rates under conditions of high ligand excess (>200), fast ligand off-rate compared to relaxation rates, saturation equilibrium conditions (saturation time = 10 s), and reasonably low protein rotational correlation time.³² The method was successfully applied to data collected outside of these limits, with ligand excess as low as 67 and saturation time 2s.^{43, 44} We confirmed the validity of this approach with our sample conditions, by comparing initial rates and T_1 -scaling at both low and high saturation times for a complex of **4** with BSA, observing no significant change to the STD profile (Supplementary Information S4).

DIRECTION experiment

The DIRECTION experiment was carried out on the same sample as used for STD measurements, using a modified pseudo-2D inversion recovery pulse sequence as described in²⁶, with interleaved on- and off-resonance irradiation of protein. On and off-resonance frequencies were the same as for the STD experiment. Irradiation was achieved using a train of 50ms Gaussian pulses during both a 3 s presaturation period and during inversion

recovery. A CPMG pulse train and WATERGATE sequence were added prior to data collection to suppress signals from protein and water respectively. The inversion recovery period T_{var} was set to 0.1 – 0.8 s to ensure initial rates and accuracy of a mono-exponential fit.²⁶ The relaxation delay was set to 5 s. Data were fit to a two-parameter single exponential equation:

$$M_{obs}=M_0 * \exp (-T_{var}/T_1^P) \quad (3)$$

where T_1^P is the T_1 observed in the presence of protein. The DIRECTION factor was obtained from the difference in the longitudinal relaxation rate (inverse of T_1) between on- and off-resonance conditions:

$$\text{DIRECTION factor} = R_1^{on} - R_1^{off} \quad (4)$$

Computational Docking

Ligand and protein setup and docking were accomplished using the OpenEye software suite (OpenEye Scientific Software, Santa Fe, NM, <http://www.eyesopen.com>). Ligand charges were assigned using *fixpka* and *molcharge* (QuacPac-1.5.0)⁴⁵. Five low energy conformations were selected for subsequent docking using *omega2-2.5.1.4*^{46–48} and refined using *szybki-1.7.0*⁴⁹. OEDocking-3.0.0^{50, 51} was used for receptor preparation and docking. The binding site was defined with *make_receptor* using a box volume of 7741 Å³ (15.33 × 21.73 × 23.67 Å) centered on the HP of PDB structure *2xra*. Inner and outer contour shapes representing the HP were 533 Å³ and 1491 Å³ respectively. Compounds were docked with the *FRED* program.^{36, 52}. Docking scores were calculated as an overall Chemgauss4 score, composed of steric, implicit solvent desolvation and hydrogen bonding terms.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by NIH grant AI122847 to MG. The authors are grateful to Prof. Dave Wemmer and Dr. Jeff Pelton at UC Berkeley for access to and technical assistance with the 800 MHz NMR. We thank Prof. Takahashi from Yokohama City University for providing the pulse sequence for the DIRECTION experiment and for helpful discussions. Molecular graphics images were produced using the UCSF Chimera package from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIH P41 RR-01081) and using OpenEye vida. The authors are grateful to OpenEye Scientific Software (<http://www.eyesopen.com>) for free academic licensing of their software.

References

1. Blumenthal R, Durell S, Viard M. J. Biol. Chem. 2012; 287:40841–40849. [PubMed: 23043104]
2. Gallo SA, Finnegan CM, Viard M, Raviv Y, Dimitrov A, Rawat SS, Puri A, Durell S, Blumenthal R. Biochim. Biophys. Acta. 2003; 1614:36–50. [PubMed: 12873764]
3. Gallo SA, Puri A, Blumenthal R. Biochemistry (Mosc). 2001; 40:12231–12236.

4. Melikyan GB, Markosyan RM, Hemmati H, Delmedico MK, Lambert DM, Cohen FS. *J. Cell Biol.* 2000; 151:413–423. [PubMed: 11038187]
5. Steger HK, Root MJ. *J. Biol. Chem.* 2006; 281:25813–25821. [PubMed: 16803885]
6. Naider F, Anglister J. *Curr. Opin. Struct. Biol.* 2009; 19:473–482. [PubMed: 19632107]
7. Sugaya M, Hartley O, Root MJ, Blauvelt A. *J. Invest. Dermatol.* 2007; 127:1436–1443. [PubMed: 17255952]
8. Otaka A, Nakamura M, Daisuke N, Kodama E, Uchiyama S, Nakamura H, Kobayashi Y, Matsuoka M, Fujii N. *Angew. Chem. Int. Ed.* 2002; 41:2938–2939.
9. Welch BD, Francis JN, Redman JS, Paul S, Weinstock MT, Reeves JD, Lie YS, Whitby FG, Eckert DM, Hill CP, Root MJ, Kay MS. *J. Virol.* 2010; 84:11235–11244. [PubMed: 20719956]
10. Welch BD, VanDemark AP, Heroux A, Hill CP, Kay MS. *Proc. Natl. Acad. Sci. U. S. A.* 2007; 104:16828–16833. [PubMed: 17942675]
11. Chan DC, Chutkowski CT, Kim PS. *Proc. Natl. Acad. Sci. USA.* 1998; 95:15613–15617. [PubMed: 9861018]
12. Chan DC, Fass D, Berger JM, Kim PS. *Cell.* 1997; 89:263–273. [PubMed: 9108481]
13. Jiang S, Tala SR, Lu H, Abo-Dya NE, Avan I, Gyanda K, Lu L, Katritzky AR, Debnath AK. *J. Med. Chem.* 2011; 54:572–579. [PubMed: 21190369]
14. Cai L, Jiang S. *Chem Med Chem.* 2010; 5:1813–1824. [PubMed: 20845360]
15. Zhou G, Sofiyev V, Kaur H, Snyder BA, Mankowski MK, Hogan PA, Ptak RG, Gochin M. *J. Med. Chem.* 2014; 57:5270–5281. [PubMed: 24856833]
16. Gochin M. *Assay Drug Dev Technol.* 2012; 10:407–416. [PubMed: 22897493]
17. Zhou G, Wu D, Snyder B, Ptak RG, Kaur H, Gochin M. *J. Med. Chem.* 2011; 54:7220–7231. [PubMed: 21928824]
18. Jubb H, Higuieruelo AP, Winter A, Blundell TL. *Trends Pharmacol. Sci.* 2012; 33:241–248. [PubMed: 22503442]
19. Rondeau, J-M., Klebe, G., Podjarny, A. *Biophysical Approaches Determining Ligand Binding to Biomolecular Targets : Detection, Measurement and Modelling.* In: Podjarny, A, Dejaegere, AP, Kieffe, B., editors. *RSC Biomolecular Sciences*, Editon edn. 2011. p. 136-155.
20. Gochin M, Whitby LR, Phillips AH, Boger DL. *J. Comput. Aided Mol. Des.* 2013; 27:569–582. [PubMed: 23893342]
21. Gochin M, Zhou G, Phillips AH. *ACS Chem. Biol.* 2011; 6:267–274. [PubMed: 21155611]
22. Dalvit C, Fogliatto G, Stewart A, Veronesi M, Stockman B. *J. Biomol. NMR.* 2001; 21:349–359. [PubMed: 11824754]
23. Dalvit C, Caronni D, Mongelli N, Veronesi M, Vulpetti A. *Curr Drug Discov Technol.* 2006; 3:115–124. [PubMed: 16925519]
24. Mayer M, Meyer B. *Angew. Chem. Int. Ed.* 1999; 38:1784–1788.
25. Meyer B, Peters T. *Angew. Chem. Int. Ed. Engl.* 2003; 42:864–890. [PubMed: 12596167]
26. Mizukoshi Y, Abe A, Takizawa T, Hanzawa H, Fukunishi Y, Shimada I, Takahashi H. *Angew. Chem. Int. Ed. Engl.* 2012; 51:1362–1365. [PubMed: 22213544]
27. Lancelot N, Piotto M, Theret I, Lesur B, Hennig P. *J. Pharm. Biomed. Anal.* 2014; 93:125–135. [PubMed: 24280017]
28. Angulo J, Nieto PM. *Eur. Biophys. J.* 2011; 40:1357–1369. [PubMed: 21947507]
29. Angulo J, Enriquez-Navas PM, Nieto PM. *Chemistry.* 2010; 16:7803–7812. [PubMed: 20496354]
30. Angulo J, Goffin SA, Gandhi D, Searcey M, Howell LA. *Chemistry.* 2016; 22:5858–5862. [PubMed: 26864212]
31. Jayalakshmi V, Krishna NR. *J. Magn. Reson.* 2002; 155:106–118. [PubMed: 11945039]
32. Kemper S, Patel MK, Errey JC, Davis BG, Jones JA, Claridge TD. *J. Magn. Reson.* 2010; 203:1–10. [PubMed: 20022272]
33. Allen WJ, Rizzo RC. *Biology.* 2012; 1:311–338. [PubMed: 23730525]
34. Walsh JD, Chu S, Zhang SQ, Gochin M. *Protein Eng Des Sel.* 2015; 28:107–116. [PubMed: 25792539]

35. Zhou G, Wu D, Hermel E, Balogh E, Gochin M. *Bioorg. Med. Chem. Lett.* 2010; 20:1500–1503. [PubMed: 20153190]
36. McGann M. J. *Comput. Aided Mol. Des.* 2012; 26:897–906. [PubMed: 22669221]
37. Yan J, Kline AD, Mo H, Shapiro MJ, Zartler ER. *J. Magn. Reson.* 2003; 163:270–276. [PubMed: 12914842]
38. Sabin C, Corti D, Buzon V, Seaman MS, Lutje Hulsik D, Hinz A, Vanzetta F, Agatic G, Silacci C, Mainetti L, Scarlatti G, Sallusto F, Weiss R, Lanzavecchia A, Weissenhorn W. *PLoS Pathog.* 2010; 6:e1001195. [PubMed: 21124990]
39. Cai L, Gochin M. *Antimicrob. Agents Chemother.* 2007; 51:2388–2395. [PubMed: 17452484]
40. Sofiyev V, Kaur H, Snyder BA, Hogan PA, Ptak RG, Hwang P, Gochin M. *Bioorg. Med. Chem.* 2017; 25:408–420. [PubMed: 27908751]
41. Hwang TL, Shaka AJ. *J. Magn. Reson. (Ser A).* 1995; 112:275–279.
42. Mayer M, Meyer B. *J. Am. Chem. Soc.* 2001; 123:6108–6117. [PubMed: 11414845]
43. Kotzler MP, Blank S, Bantleon FI, Spillner E, Meyer B. *Biochim. Biophys. Acta.* 2012; 1820:1915–1925. [PubMed: 22982178]
44. Lawrence AD, Frank S, Newnham S, Lee MJ, Brown IR, Xue WF, Rowe ML, Mulvihill DP, Prentice MB, Howard MJ, Warren MJ. *ACS Synth Biol.* 2014; 3:454–465. [PubMed: 24933391]
45. QUACPAC-1.5.0. OpenEye Scientific Software, Santa Fe, NM. <http://www.eyesopen.com>, Editon edn.
46. Hawkins PC, Nicholls A. *J Chem Inf Model.* 2012; 52:2919–2936. [PubMed: 23082786]
47. Hawkins PC, Skillman AG, Warren GL, Ellingson BA, Stahl MT. *J Chem Inf Model.* 2010; 50:572–584. [PubMed: 20235588]
48. Hawkins, PCD., Skillman, AG., Warren, GL., Ellingson, BA., Stahl, MT. OMEGA2.5.1.4. OpenEye Scientific Software, Santa Fe, NM. <http://www.eyesopen.com>, Editon edn.
49. SZYBKI-1.7.0. OpenEye Scientific Software, Santa Fe, NM. <http://www.eyesopen.com>, Editon edn.
50. Kelley BP, Brown SP, Warren GL, Muchmore SW. *J Chem Inf Model.* 2015; 55:1771–1780. [PubMed: 26151876]
51. OEDOCKING-3.0.0. OpenEye Scientific Software, Santa Fe, NM. <http://www.eyesopen.com>, Editon edn.
52. McGann M. *J Chem Inf Model.* 2011; 51:578–596. [PubMed: 21323318]

figure 1. depiction of reverse hairpin proteins used in the study. helices are represented by ovals connected by the short loop l4. chr residues are in green and nhr residues in red, and the location of the hp indicated as a yellow oval. the n- and c-termini of the protein are labeled, as are the lengths of each subdomain. three reverse hairpins with different lengths of the chr domain are indicated. only one monomer of the trimer of hairpins that forms along the nhr axis is shown.

A. 1D ^1H NMR of **4** B. WaterLOGSY NMR spectra of **4** **figure 2.**

aromatic region of nmr spectrum of **4** at 400 mhz. a. 1d spectrum of 1 mm **4** in aqueous phosphate buffer (50 mm, ph 7.4) containing 10% d₂o and 2% dms₂-d₆; b. waterlogsy spectra of 1mm **4** in buffer or 0.5 mm **4** in the presence of three reverse hairpin gp41 constructs (at 10 μ m), as indicated. the peak at 8.46 ppm is formate.

figure 3.

clusters of poses of 4 obtained by computational docking in the hp of pdb structure 2xra. two nhr chains form the hp. residue numbering is for hxb2 with a' to distinguish residues from the second protein chain. hydrogen atoms are omitted for clarity. carbon atoms of residues at the base of the hp are shown in green, and those surrounding the pocket in yellow. oxygen and nitrogen atoms are colored in red and blue respectively. in the first panel, carbon atoms of the three poses 1, 2 and 5 of cluster 1 are shown in blue, purple and green respectively. in all other clusters, carbon atoms of the ligand are gray.

figure 4.
800 mhz nmr spectra of 4. a. 1d aromatic proton nmr of 1 mm 4 with assignments. b. std difference spectrum of 1 mm 4 in the presence of 10 μm c28(l4)n50, at a saturation time of 3s. the chemical structure and relative qstd-af values (equation 2) are shown as inserts, color-coded for weak, medium and strong. all spectra were obtained in 50 mm phosphate buffer, ph 7.4, containing 2% dms0-d6 and 100 μm tsp.

figure 5. ligand epitope profiles from a. std and b. direction experiments. in a, the error bars are the standard deviation of qstd values obtained using t1's of ligand in the absence or presence of protein. in b, the error bars were calculated from the sum of errors of fitting the on – and off-resonance relaxation rates using equation 3.

figure 6.

identification of the binding mode of 4 in the hp of gp41 from std profile pattern. a. observed std factors (equation 2); b–f. calculated sums of intermolecular protein – ligand $1/r^6$ distances for all non-exchangeable protein protons surrounding each ligand proton; b. calculated for pose 1 as representative of cluster 1. c. calculated for pose 3 as representative of cluster 2; d, f. calculated for poses 6 and 9 as representative of cluster 3; e. calculated for pose 7 as representative of cluster 4. y values in b–f are $1000 \cdot s(1/r^6)$ for all non-exchangeable ligand – protein proton pairs within $r \sim 6.5$ Å.

figure 7. relative protein-ligand noe values calculated for each observable proton position in cluster 1. a–c. calculated noe's for poses 1, 2 and 5, respectively, color-coded as for the ligands shown in figure 3a. d. average of the three plots in a–c, with error bars showing the standard deviation. the scale on the y-axis is $1000 \cdot \sigma(1/r^6)$ for all non-exchangeable ligand – protein proton pairs within $r \sim 6.5$ Å.

figure 8.

pose 1 of cluster 1 in the hp of 2xra. the pose agreed with available std data, which is shown superimposed on the structure as circles representing weak (blue), medium (orange) and strong (red) std intensities at each ligand proton position. absence of a circle indicates no data was available at that position due to overlap. residues from two nhr chains forming the base of the pocket and lining the pocket are labeled. a hydrogen bond predicted to form between the ligand carboxylate and pocket arg is shown as a green dashed line. possible clashes between ligand and pocket residues are shown as yellow dashed lines.

Scheme 1.

Reagents and Conditions: (a) $\text{Pd}[\text{PPh}_3]_4$, K_2CO_3 , THF, Argon, 80°C . (b) NaH/DMF, Methyl-3-(bromomethyl)benzoate; (c) 2.0 M NaOH; 1.0M HCl.