Published in final edited form as:

Bioorg Med Chem Lett. 2013 September 15; 23(18): . doi:10.1016/j.bmcl.2013.07.026.

Identification of Fragments Targeting an Alternative Pocket on HIV-1 gp41 by NMR Screening and Similarity Searching

Shidong Chu^a and Miriam Gochin^{a,b,*}

^aDepartment of Basic Sciences, College of Osteopathic Medicine, Touro University California, Vallejo, CA 94592

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

Abstract

The HIV-1 envelope glycoprotein gp41 fusion intermediate is a promising drug target for inhibiting viral entry. However, drug development has been impeded by challenges inherent in mediating the underlying protein-protein interaction. Here we report on the identification of fragments that bind to a C-terminal sub-pocket adjacent to the well-known hydrophobic pocket on the NHR coiled coil. Using a specifically designed assay and ligand-based NMR screening of a fragment library, we identified a thioenylaminopyrazole compound with a dissociation constant of $\sim 500\mu\text{M}$. Interaction with the C-terminal sub-pocket was confirmed by paramagnetic relaxation enhancement NMR experiments, which also yielded the binding mode. Shape-based similarity searching detected additional phenylpyrazole and phenyltriazole fragments within the library, enriching the hit rate over random screening, and revealing molecular features required for activity. Discovery of the novel scaffolds and binding mechanism suggests avenues for extending the interaction surface and improving the potency of a hydrophobic pocket binding inhibitor.

© 2013 Elsevier Ltd. All rights reserved.

Corresponding Author: Miriam Gochin, Touro University California, 1310 Club Drive, Mare Island, Vallejo CA 94122, Tel: (707) 638-5463; miriam.gochin@tu.edu.

Supplementary Data: Supplementary data includes detailed experimental procedures for fragment screening, NMR and computational analysis, and additional figures.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Keywords

HIV-1; gp41; fragment library; NMR; WaterLOGSY; ROCS similarity searching

Low molecular weight fusion inhibitors targeting the highly conserved transmembrane glycoprotein-41 (gp41) of HIV-1 would be valuable additions to the anti-HIV therapeutic arsenal, due to their complementary mode of action compared to current oral antiretroviral medications.¹ However, it has been difficult to find potent non-peptide inhibitors against the gp41 intramolecular protein – protein interaction central to HIV fusion.^{2, 3} C-terminal heptad repeat (CHR) helices of gp41 fold onto an N-heptad repeat (NHR) helical trimer during the conformational transition that accompanies fusion. Examining specific interaction sites and inhibitor binding modes on the NHR trimer has been hindered by its poor solubility and by the inability to crystallize gp41 with small molecules.

Most studies investigating low molecular weight fusion inhibitors have been directed at a deep hydrophobic cavity on the NHR trimer⁴, yielding compounds with μM potency^{2, 5-9}. Increasing potency has been a challenge and may represent an inability to mediate the extended protein – protein interface between NHR and CHR with a small molecule limited to the hydrophobic cavity. C-peptides extended downstream from the hydrophobic cavity – binding residues are strongly helical and become consistently more potent with length.¹⁰ Upstream extension of C-peptides by one to two heptad repeats also improves potency and 6 helix bundle stability, although there is an entropic penalty due to flexibility and low helicity of C-peptide in this region.¹¹⁻¹³

Fragment-based drug design has been increasingly applied as a promising strategy for mediating protein – protein interactions.^{14, 15} In the case of gp41, the long extended interface contains several sub-pockets that could in principle be occupied by fragments (< 300da) engaged in high quality interactions. These could be linked chemically to form potential lead compounds for drug design. In this study, we have used gp41 constructs designed to detect fragment binding to the NHR region C-terminal to the hydrophobic cavity, in order to exploit the stabilizing interactions formed by the C-peptide segment upstream of the cavity binding domain. The relatively high polarity of NHR – CHR residues in this region and possibility of forming hydrogen bond or buried salt-bridge interactions suggest it may be an attractive site for development of small drug-like molecules. Furthermore, these molecules bind immediately adjacent to the hydrophobic cavity, leading to the possibility of linking and extending hydrophobic pocket binding inhibitors. WaterLOGSY (Water-Ligand Observed via Gradient Spectroscopy) NMR experiments¹⁶ have led to the discovery of several aryl and amino-group substituted pyrazoles and triazoles which bind with mid μM affinity to the C-terminal cavity. The binding mode of one of the derivatives has been evaluated using paramagnetic relaxation enhancement (PRE) NMR spectroscopy.^{17, 18}

A few crystal structures include the C terminal sub-pocket flanked by residues 579-RILAVEYLKDQ-590 at the C-terminal end of the NHR helical region (3vgx, 3p7k)¹³. The apposing CHR residues are not well defined in crystal structures (1env, 3vgx), but a homology modeled structure derived from the solution NMR structure of Simian Immunodeficiency Virus (SIV) gp41 (1if3)¹⁹ has the side chain of K617 projecting into this pocket. Screening for binding in the sub-pocket was accomplished using a pair of peptides env3.0 and C18-e3.0 representing the NHR and CHR domains respectively, as shown in Figure 1. These are helical segments from the hairpin structure of the gp41 extracellular domain. Env3.0 was stabilized as a trimeric coiled coil structure by octahedral coordination of Fe(II) with three N-terminal bidentate bipyridyl (bpy) groups.^{11, 20} C18-e3.0 contains the

hydrophobic pocket binding domain WeeWDkeI, and has 0.5 μ M affinity to Fe(env3.0)₃.¹¹ Association of the two peptides resulted in the N-terminal cysteine residue of C18-e3.0 situated at the C-terminal edge of the hydrophobic pocket (Figure 1B, D). Spin-labeling at this cysteine was used to provide a probe for PRE determination of bound ligand orientation in the sub-pocket.

A 500-member fragment library was purchased from Chembridge, containing compounds which conformed to the Rule-of-Three.²¹ Fragments were selected to contain at least one aromatic group and some polar functionality. Fragments were clustered into groups of ten at a concentration of 1mM each, and WaterLOGSY experiments were performed in the absence and presence of 10 μ M Fe(env3.0)₃ (30 μ M in binding sites) and 30 μ M C18-e3.0. Figure 2 shows an example of this experiment on a ten fragment mixture, which resulted in the identification of fragment **1a**, 4-methyl-5-thioenyl-3-aminopyrazole, as a hit by resonance inversion.¹⁶ A second fragment, 2-(benzylthio)ethanamine, was also identified as a hit, although it is not progressed further here due to its flexible structure and the absence of similar compounds in the library for SAR exploration (see below). **1a** (inset in Figure 2) and 2-(benzylthio)ethanamine are reminiscent of a lysine side chain in hydrophobicity and terminal charge. Recalling that a lysine side chain potentially occupies the sub-pocket suggests there may be a similar binding mechanism for the fragments.

Fragment NMR screening was considerably hindered and prolonged by practical complications including selection of fragments for clustering to avoid NMR signal overlap and issues of low solubility. We noted that some fragments when clustered caused aggregation and lead to artifacts in screening. False positives were detected by inversion of one or more peaks in the absence of protein and those clusters were discarded. Remaining fragments were reclustered without the offending aggregators. Fragments were also screened individually if there was significant observed overlap of resonances.

Due to the low throughput nature of the NMR screening, we conducted a shape and chemical similarity search on the library to identify compounds in the library that were similar to **1a** and accelerate their testing. Fragment overlap was calculated using the 3D shape and chemistry based alignment method ROCS (Rapid Overlay of Chemical Structures) from OpenEye.^{22, 23} We identified 29 compounds, with a ROCS Tanimoto-Combo (RT) score > 1.18. The RT score is a combination of a molecular shape similarity coefficient and a color force field based on hydrogen bond donors and acceptors, anions, cations, hydrophobes and rings, and has a value ≤ 2 . Below RT=1.18, fragment structure changed significantly. Of the thirty compounds selected by ROCS (including **1a**), seventeen compounds tested positive, ten negative, two lacked non-exchangeable protons and one was not soluble enough for NMR studies. Thus the hit-rate for the targeted sub-library was at least 57%, which is significantly enhanced over the random screening hit rate of 6% that we observed studying 320 fragments by NMR without considering shape or chemical similarity. The selection of this sub-library does not rule out the possibility that chemically different fragment hits could be identified by continued experimental screening.

The chemical similarity of fragment hits is an indication of a specific interaction in the C-terminal pocket and demonstrates the druggability of the pocket. Figure 3 shows the structures of the NMR-observable ROCS-selected fragments, color-coded for activity.

Three additional compounds are included in the figure: compound **1e** which was purchased subsequently as a chemically similar fragment for ¹⁹F detection, and compounds **1f** and **4d**, which have RT=0.987 and RT=0.774 respectively against the query **1a**, but which were (correctly) identified by eye as possible hits based on their molecular properties (see below). Clear features contributing to activity were apparent from the structures, even with this

small sampling of compound space. All except six of the compounds contained a 5-membered pyrazole or triazole ring in which an amino group substituent was essential for activity (scaffolds **1** – **4**). The heterocyclic ring also had one or two additional non-polar substituents, including a strong hydrophobe, typically an aryl group. Activity was abrogated if the amino group was replaced by hydrogen, thioxo, carboxylic acid or carbohydrazine, or if one of the non-polar substituents was replaced with a second amino group (scaffolds **5** – **10**). Of the remaining six compounds, three were similar to the amino-substituted pyrazoles, but with a pyrimidine ring instead. Two of these **11a,b** were hits; the third **12** which sported an additional thioxo was inactive. A fourth triazine ring compound **13** missing the amino or aryl substituents was inactive. The remaining two compounds **14** and **15**, both active, were chemically more distant from the query structure, ranking at the bottom of the list of 30 compounds by RT score. Compound **15** (RT=1.183) appears to represent a scaffold hop, and could be used as a new query to explore a different region of compound space for fragments and compounds targeting the pocket.

The electrostatic potentials of the 27 NMR-observable fragments aligned by ROCS were calculated using the tool EON from OpenEye, which reports an overall electrostatic Tanimoto (ET) score as a combination of the Poisson-Boltzmann and EON Shape components of the electrostatic potential.^{22, 24} Both shape and electrostatics are important ligand-based parameters for predicting actives, and the ROCS and EON scores were plotted against each other in Figure 4. Compounds with an RT score above 1.25 and an ET score above 0.93 were reliably predicted as actives. RT scores between 1.18 and 1.25 were not predictive. All compounds except **15** with an ET score < 0.9 were inactive.

WaterLOGSY NMR experiments did not detect aggregation of **1a** even at much higher concentration (20mM) than that used in screening, confirming a specific association with the peptide pair Fe(env3.0)₃ / C18-e3.0. Using a fluorescence-based binding assay developed previously in our group²⁵, we have also identified several fragments from the same library that bind to the hydrophobic pocket of gp41. **1a** tested negative in the hydrophobic pocket – specific fluorescence assay (Figure S1). Additionally, a 7μM hydrophobic pocket binding fragment did not show reversed peaks in WaterLOGSY spectra in the presence of Fe(env3.0)₃ / C18-e3.0 (Figure S2), confirming that C18-e3.0 was not displaced by the HP-binding fragment. These results confirm that the fragment hits identified in this work are not binding to the hydrophobic pocket of gp41.

We used paramagnetic relaxation to probe the orientation of **1a** bound in the sub-pocket.^{17, 18} C18-e3.0 was labeled with MTSL at the N-terminal cysteine, exposing the binding site on Fe(env3.0)₃ to a transverse relaxation gradient (Figure 1B, 1D). Relaxation rates of the four observable protons of **1a** were determined for various concentrations of Fe(env3.0)₃ and C18-e3.0, ranging from 10 – 40 μM. Differences in relaxation rates obtained with spin-labeled vs. unlabeled C18-e3.0 were converted into PRE values for the bound form of **1a**, using K_D = 500μM measured with the WaterLOGSY experiment.¹⁶ The PRE's were converted into distances using the Solomon-Bloembergen equation²⁶ and included in XPLOR calculations using the NOE square well function. Details of the PRE calculation are found in references^{17, 18} and in the Supplementary Data.

The four PRE-derived distance constraints were used to examine 9 docked poses (Figure S3) of **1a** obtained using AutoDock Vina²⁷. Protein side chains, including MTSL, were allowed to move in a minimization protocol that evaluated fixed ligand poses for agreement with the NOE's. Protocol details are provided in the Supplementary Data. Figure 5 shows the results obtained for each of the nine poses and for four correlation times (τ_c) ranging from 8 – 14ns.^{17, 18} Figure 5A shows that only pose 1 agreed with the data, having low NOE RMSD (0.008 Å at τ_c = 10ns) and no distance violations > 0.1Å at τ_c = 10ns or higher. Minimal side

chain rearrangements occurred during minimization (Supplementary Figure S4). The XPLOR-minimized structure with pose 1 is shown in Figure 5B. The amino group makes a putative hydrogen bond to the backbone carbonyl of Arg579. Although there are undoubtedly variations in the exact binding mode of the active fragments listed in Figure 3, the necessity of an amino group and the loss of activity associated with its removal or substitution of a carboxy or carbonyl moiety is explained by this interaction. Furthermore, the requirement for hydrophobic substituents on the pyrazole or triazole ring is explained by the hydrophobic interactions they make with pocket residues.

In summary, a limited focused NMR screening campaign guided by cheminformatic analysis provided a novel series of fragment binders with a distinct mode of interaction with gp41. ROCS similarity searching resulted in more efficient library mining and identification of molecular properties required for activity. The common molecular features of the hits were amino- and aryl-substituted pyrazole or triazole rings. The NMR-validated pose of representative hit **1a** revealed specific polar and hydrophobic interactions within a small C-terminal pocket adjacent to the hydrophobic pocket. The binding pose partially emulates a lysine residue from the loop region that is predicted to reside in this cavity (Supplementary Figure S5).¹⁹ **1a** specificity and ligand efficiency ($G/N_{\text{heavy atoms}} = 0.38$) are indicative of the ligandability of the pocket.¹⁴ Discovering fragments is an encouraging step in the complex and challenging process of mediating the protein – protein interaction.^{15, 28, 29} Our results provide for the first time a series of fragment binders to a relatively shallow sub-site close to the deep hydrophobic pocket of gp41, which may lead to cooperative binding when the fragments are tethered with hydrophobic pocket binders. The results reported here can be productively utilized with further SAR, structural biology and medicinal chemistry for gp41 fusion inhibitor development.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We gratefully acknowledge the financial support of the National Institutes of Health (GM087998 to MG). 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). We thank OpenEye Inc. Scientific Software for providing an academic license. We thank Prof. Marc Anderson for assistance with scripts to run OpenEye software, and in selection of fragments to include in the library.

References

1. Eckert DM, Kim PS. *Annual Rev Biochem.* 2001; 70:777. [PubMed: 11395423]
2. Cai L, Jiang S. *ChemMedChem.* 2010; 5:1813. [PubMed: 20845360]
3. Naider F, Anglister J. *Curr Opin Struct Biol.* 2009; 19:473. [PubMed: 19632107]
4. Chan DC, Chutkowski CT, Kim PS. *Proceedings of the National Academy of Sciences of the United States of America.* 1998; 95:15613. [PubMed: 9861018]
5. Zhou G, Wu D, Hermel E, Balogh E, Gochin M. *Bioorg Med Chem Lett.* 2010; 20:1500. [PubMed: 20153190]
6. Zhou G, Wu D, Snyder B, Ptak RG, Kaur H, Gochin M. *J Med Chem.* 2011; 54:7220. [PubMed: 21928824]
7. Gochin M, Zhou G. *Curr Topics Med Chem.* 2011; 11:3022.
8. Whitby LR, Boyle KE, Cai L, Yu X, Gochin M, Boger DL. *Bioorg Med Chem Lett.* 2012; 22:2861. [PubMed: 22424973]
9. Wang Y, Lu H, Zhu Q, Jiang S, Liao Y. *Bioorg Med Chem Lett.* 2010; 20:189. [PubMed: 19932616]

10. Gochin M. *Assay Drug Dev Technol.* 2012; 10:407. [PubMed: 22897493]
11. Gochin M, Cai L. *J Med Chem.* 2009; 52:4338. [PubMed: 19534533]
12. He Y, Cheng J, Li J, Qi Z, Lu H, Dong M, Jiang S, Dai Q. *J Virol.* 2008; 82:6349. [PubMed: 18417584]
13. Chong H, Yao X, Qiu Z, Qin B, Han R, Waltersperger S, Wang M, Cui S, He Y. *J Biol Chem.* 2012; 287:20281. [PubMed: 22511760]
14. Bower JF, Pannifer A. *Current Pharmaceutical Design.* 2012; 18:4685. [PubMed: 22650253]
15. Winter A, Higuero AP, Marsh M, Sigurdardottir A, Pitt WR, Blundell TL. *Quarterly Reviews of Biophysics.* 2012; 45:383. [PubMed: 22971516]
16. Dalvit C, Fogliatto G, Stewart A, Veronesi M, Stockman B. *J Biomol NMR.* 2001; 21:349. [PubMed: 11824754]
17. Balogh E, Wu D, Zhou G, Gochin M. *J Am Chem Soc.* 2009; 131:2821. [PubMed: 19206471]
18. Gochin M, Zhou G, Phillips AH. *ACS Chem Biol.* 2010; 6:267. [PubMed: 21155611]
19. Caffrey M. *Biochim et Biophys Acta.* 2001; 1536:116.
20. Gochin M, Guy RK, Case MA. *Angew Chem Int Ed.* 2003; 42:5325.
21. Congreve M, Carr R, Murray C, Jhoti H. *Drug Discov Today.* 2003; 8:876. [PubMed: 14554012]
22. Nicholls A, Grant JA. *J Comput Aided Mol Des.* 2005; 19:661. [PubMed: 16328855]
23. Hawkins PC, Skillman AG, Nicholls A. *J Med Chem.* 2007; 50:74. [PubMed: 17201411]
24. Nicholls A, MacCuish NE, MacCuish JD. *J Comput Aided Mol Des.* 2004; 18:451. [PubMed: 15729846]
25. Cai L, Gochin M. *Antimicrob Agents Chemother.* 2007; 51:2388. [PubMed: 17452484]
26. Solomon I, Bloembergen N. *J Chem Phys.* 1956; 25:261.
27. Trott O, Olson AJ. *J Comput Chem.* 2010; 31:455. [PubMed: 19499576]
28. Wells JA, McClendon CL. *Nature.* 2007; 450:1001. [PubMed: 18075579]
29. Makley LN, Gestwicki JE. *Chem Biol Drug Des.* 2013; 81:22. [PubMed: 23253128]

Abbreviations

gp41	glycoprotein-41
ROCS	Rapid Overlay of Chemical Structures
WaterLOGSY	Water-Ligand Observed via Gradient Spectroscopy
PRE	Paramagnetic Relaxation Enhancement

Figure 1.

Gp41 peptides for detection of C-terminal pocket binders. A. Schematic representation of the extracellular hairpin of gp41, with helical residues numbered and shown as rectangles; B. NHR (env3.0) and CHR (C18-e3.0) peptides selected for exposing the C-terminal pocket (indicated by the arrow); C. Amino acid sequences of env3.0 and C18-e3.0. Env3.0 includes 4 residues preceding and 3 residues following the gp41 sequence; D. Structural model of C18-e3.0 bound to Fe(env3.0)₃ showing hydrophobic pocket binding residues W628, W631 and I635.

Figure 2.

NMR screening of a ten-fragment mixture, with each fragment at 1mM. A. WaterLOGSY and B. 1D ^1H NMR spectra in the presence of 10 μM Fe(env3.0) $_3$ (30 μM in binding sites) and 30 μM C18-e3.0. The spectra were recorded in 5% DMSO- d_6 /10% D $_2$ O/85% H $_2$ O in deuterated 25mM Tris-acetate buffer (pH 7.5). Water suppression was achieved using excitation sculpting with gradients. Dashed lines highlight resonances of **1a** that are inverted in the WaterLOGSY spectrum. The inverted peak at 7.4 ppm indicated by a * is a second hit in this mixture (2-(benzylthio)ethanamine, 5 overlapping protons).

Figure 3.

Pharmacophore classification of fragments selected by ROCS searching with query **1a** and identified by NMR screening as actives (blue) or inactive (magenta). The fragments were classified on the basis of chemical structure.

Figure 4. Plot of ROCS RT score vs EON ET score for the 26 fragments generated by similarity searching using fragment **1a** as the ROCS query. Active and inactive compounds are shown as closed and open circles respectively. The grey shaded area corresponds to regions of RT and ET in which both actives and inactives were found.

Figure 5. NMR analysis of docked poses of **1a**. A. NOE violations $> 0.1 \text{ \AA}$ and NOE RMSD are shown for correlation times 8, 10, 12, 14 ns for each AutoDock Vina pose 1 – 9. In brackets are the PRE RMSD values in s^{-1} at $\tau_c = 10 \text{ ns}$. B. Docked pose 1 validated by PRE data is shown in the C-terminal pocket of pdb structure 3p7k, with MTSL-labeled C18-e3.0 docked into the hydrophobic pocket. The position of MTSL was determined by minimization against the PRE-derived NOE constraints. Receptor atoms within $3 \text{ \AA}$ or $4 \text{ \AA}$ of **1a** are colored in orange and yellow respectively.