DISSERTATION

ENGINEERING AND EVOLVING HELICAL PROTEINS THAT IMPROVE IN VIVO STABILITY AND INHIBIT ENTRY OF ENFUVIRTIDE-RESISTANT HIV-1

Submitted by

Susanne N. Walker

Department of Chemistry

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Spring 2019

Doctoral Committee:

Advisor: Alan Kennan

Tingting Yao
Andrew McNally
Robert Paton

ABSTRACT

ENGINEERING AND EVOLVING HELICAL PROTEINS THAT IMPROVE IN VIVO STABILITY AND INHIBIT ENTRY OF ENFUVIRTIDE-RESISTANT HIV-1

Methods for the stabilization of well-defined helical peptide drugs and basic research tools have received considerable attention in the last decade. Enfuvirtide is a 36-residue chemically synthesized helical peptide that targets the viral gp41 protein and inhibits viral membrane fusion. Enfuvirtide-resistant HIV, however, has been prolific, and this peptide therapy requires daily injection due to proteolytic degradation.

In this dissertation I have developed a method for stabilizing helical peptide therapeutics termed helix-grafted display proteins. These consist of the HIV-1 gp41 C-peptide helix grafted onto Pleckstrin Homology domains. Some of these earlier protein biologics inhibit HIV-1 entry with modest and variable potencies (IC_{50} 190 nM - >1 μM). After optimization of the scaffold and the helix, our designer peptide therapeutic potently inhibited HIV-1 entry in a live-virus assay (IC_{50} 1.9-12.4 nM). Sequence optimization of solvent-exposed helical residues using yeast display as a screening method led to improved biologics with enhanced protein expression in *Escherichia coli* (*E. coli*, a common bio-expression host), with no appreciable change in viral membrane fusion suppression. Optimized proteins suppress the viral entry of a clinically-relevant double mutant of HIV-1 that is gp41 C-peptide sensitive and Enfuvirtide-resistant. Protein fusions engineered for serum-stability also potently inhibit HIV-1 entry.

ACKNOWLEDGMENTS

For the immense work discussed in these pages, I have my greatest support group to thank, my family and close friends. Completing a doctorate degree is no easy accomplishment. It comes with diligence and sacrifice. The people in my life that dealt with this sacrifice of personal relationships, time, and stress were also the people cheering me on and standing by when times were toughest. This kind of support is rare and genuine, and it speaks volumes of their humanity and generosity. I want to thank each and every one of these people properly.

Specifically, I have my mother and father to thank for their relentless support and their initial inspiration for my love of the sciences. The musings of physics, life and the universe came from conversations with them growing up and it sparked a fire of curiosity for me which I am so grateful for. I also have my close friends to thank for this work. These girls are as close to sisters as my own dear sister, and despite the minimal time I had to offer as a sister or a friend, these women always showed up and inspired me to work harder and push through the difficulties that life brings. I am so thankful for these sisters; Krissy Walker, Jodie Sims, Katherine Webb, Erin O'Grady, and Kaci Sintek. I need to thank another person who has been supportive through the biggest challenges graduate school had to offer, Austin Cale. This career choice is difficult for friendships, but possibly even more so for relationships. Austin showed me what it was to be the best version of myself through challenges and he pushed me to face fears optimistically. He made many sacrifices to allow my scientific pursuits to come to fruition and he did it with patience and grace. This work acknowledges the support and sacrifice of this genuine group of people and I am so ever thankful for them.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

Protein Engineering and Evolution Approaches to Evaluate and Inhibit Helix-Dependent Protein-Protein Interactions

1.1 Introduction

The α-helix is the most prominent secondary structure in protein folds.[1] This common motif is found in protein-protein interactions (PPIs) involved in virtually all aspects of cellular function and signaling.[2] Due to the ubiquitous nature of this secondary structure, it has been of immense importance and focus for scientific research for the last century and probing or inhibiting these helical interactions has a high pharmacological relevance.[3, 4]

Diseased states involving a helical ligand encompass a variety of cancer receptors, either intracellular or extracellular, as well as viral membrane fusion of some of the most lethal viruses plaguing humanity. Human epidermal growth factor receptor 2 (HER2) is an extracellular target overexpressed on breast cancer cells which has become a target of particular interest in the cancer community.[5] The tumor suppressor p53 has been found to contain alterations in a large majority of cancers preventing functional activation which signals for apoptosis of cancer cells. Turning on this pathway could be a potential route for eradicating cancer cells.[6-8] HIV, Lassa Virus, Ebola virus all contain viral envelope proteins which fold into a 6-helix bundle, promoting membrane fusion and cell entry of the viral cargo. Inhibition of membrane fusion not only inhibits viral entry, but can prevent latent reservoirs of retroviruses from forming.[9-11] This review will highlight the advancements made to target helical interfaces of critical PPIs.

1.2 Bacterial Expression of Peptides

The perhaps most obvious solution to probing and inhibiting protein-protein interactions involving an α-helix is to mimic the helical ligand involved in the PPI with a peptide.[2, 12] While peptides can be manufactured through synthetic routes using solid-phase peptide synthesis, many efforts have been advanced to effectively express and purify a peptide in *Escherichia coli* (*E. coli*).[13, 14] This significantly lowers the cost of production and improves the yield, especially in the case of extended helical peptides (greater than 20 amino acids). Expressing a peptide by itself in *E. coli* can be problematic because many peptides lack solubility in a sequence-dependent manner and, in the absence of a stabilizing tertiary structure, are unstable and partially unfolded in solution making them particularly susceptible to peptidases and proteolytic degradation. Recombinant protein fusion approaches overcome the stability and solubility challenges, allowing for protein expression in a bacterial system and cost-effective purification of the peptide.[14]

Successful recombinant protein fusions bestow enhanced stability, and often solubility to the peptide. The fusion protein itself must exhibit hyperstability and high expression in the *E. coli* host to overcome the tendency for the insoluble peptide-of-interest to aggregate in inclusion bodies. The fusion must also contain an affinity tag for purification and a sequence-specific method for cleaving the protein fusion post-purification to separate the peptide (Figure 1.2A).[13-16] Proteins such as NusA and thioredoxin (TRX) exhibit the solubility properties necessary for peptide expression but must include a separate affinity tag fusion for purification and a protease cleavage site (Figure 1.2B).[17-19] Small ubiquitin-like modifier (SUMO), and ubiquitin (Ub) contain a site-specific recognition for enzymatic cleavage of the protein fusion from the peptide-of-interest, but still require an affinity tag for purification (Figure 1.2C).[20-24] Glutathione S-transferase (GST), maltose binding protein (MBP) and protein-A function as fusion proteins and

affinity tags, but still require a protease cleavage site to separate the peptide from the fusion post-purification (Figure 1.2D).[15, 25-28]

In a somewhat unique approach, a new emerging class of fusion proteins sequesters the protein products in inclusion bodies. This protects the peptides from proteolytic degradation and allows for easy separation through centrifugation of cell lysates. However, many of these fusions, including bacterial ketosteroid isomerase (KSI), the autoprotease Npro of classical swine fever virus and the elastin-like polypeptide (ELP), require chemical cleavage of the fusion from the peptide and often involve a refolding procedure for a functional peptide (Figure 1.2E).[29-31] In a recent study, Zhao et. al. combined the self-assembling peptide, ELK16, with a self-cleavable *Mxe* GyrA intein tag to produce a fusion that protects the peptide-of-interest in inclusion bodies during expression and centrifugal separation and induces cleavage of the fusion in the presence of dithiothreitol (DTT). If the peptide is soluble, the pure product will be released to the soluble fraction post-cleavage.[32]

Whether produced through chemical synthesis or bacterial expression, peptides resemble the native binding interaction and can be used to probe PPIs or inhibit their interactions. However, peptides are not considered ideal candidates for therapeutic design. In solution, they are unstable and partially unfolded leaving an entropic cost associated with binding and are once again susceptible to proteolytic degradation and rapid recycling *in vivo*. Counteracting these drawbacks has been at the forefront of research for therapeutics involving a helical ligand.

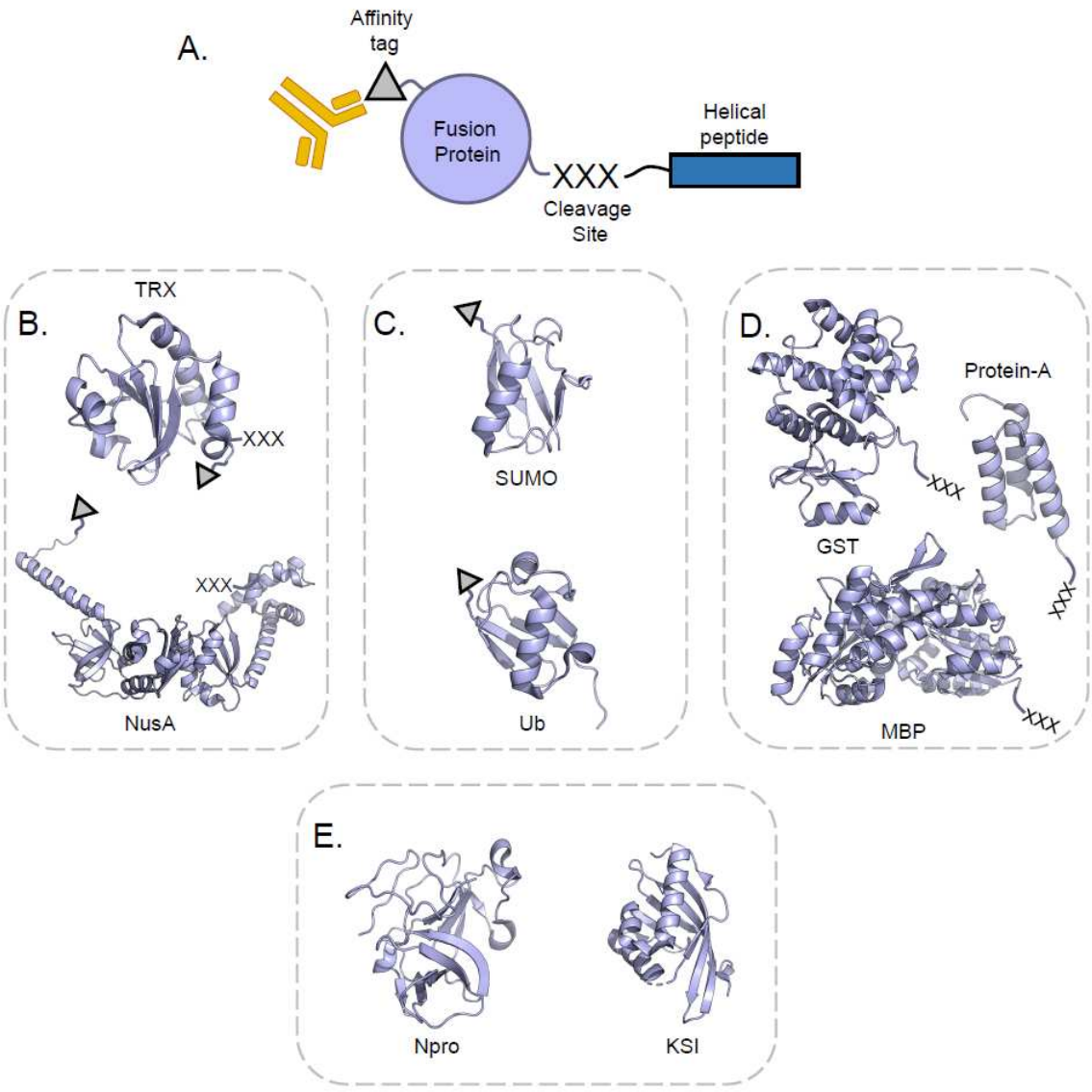


Figure 1.2 (A) Peptide expression schematic. Properties necessary for expressing a functional peptide (blue) using protein fusions requires: highly stable fusion protein to keep peptide in the soluble lysate (lavender), an affinity tag for purification (grey), and a protease cleavage site for peptide isolation (black). (B) Fusion proteins NusA and TRX which require an affinity tag and a protease cleavage site (PDB: 5LM9 and 1T00). (C) Fusion proteins SUMO and Ub which only require an affinity tag as the proteins act as a self-cleavable fusion (PDB: 1WM3 and 4MDK). (D) Protein fusions GST, MBP, and Protein-A which only require a protease cleavage site as the fusion protein acts as an affinity tag (PDB: 5LCZ, 1NMU and 4WWI). (E) Fusion proteins Npro and KSI which function by protecting peptide in insoluble fraction and can be isolated from the cell inclusion bodies with chemical cleavage (PDB: 3ZFN and 3VSY).

1.3 Chemical Approaches to Helical Stabilization

Small molecules have been an attractive solution for developing drug leads over the last century. Typically, small molecule approaches to targeting protein-protein interactions have been to mimic substrates or transition state analogues that fit within a well-defined hydrophobic pocket (Figure 1.3A).[33] Protein-protein interactions that involve a helical ligand generally do not fit within this canonical small molecule model and are challenging to target due to the large surface area involved in the binding interface. Researchers have come up with alternative chemical approaches to disrupt protein-protein interactions involving an α-helix through the synthesis of peptide mimetics; either functional mimetics that do not resemble a helical structure but still bind the identical interface or helical mimetics that functionally and structurally resemble a helical peptide.[34]

There have been some advancements of functional mimetics that bind the chemical space occupied by a specific α-helix, although this strategy has yielded few developments towards a general approach to target helical binding pockets. Hoffmann–La Roche, Shaomeng Wang group and Johnson and Johnson Pharmaceuticals generated small molecule drugs that inhibit the p53/MDM2 interaction. These inhibitors were found using high-throughput screening or virtual screening and include tetra-substituted imidazoles (Nutlins), spirooxindoles and benzodiazepinediones (Figure 1.3B).[35-37]

Helical mimetics that both functionally and physically resemble a peptide have shown vast improvements in disrupting disease-relevant PPIs as well as general engineering strategies for targeting helical interfaces. Constrained peptides, foldamers and proteomimetic-derived ligands bind to the helical interface and can improve stability compared to peptides, prevent proteolytic degradation, and target a larger surface area than the typical small molecule drugs.[38] Constrained

peptides comprise of a number of different motifs including cyclic peptides at their simplest form to bicyclic peptides and stapled peptides. Biophysical and Analytical Technologies at Aileron developed a stapled peptide that also targeted the p53/MDM2 PPI. The phase II drug candidate, ALRN-6924, binds to both MDM2 and MDMX by resembling the helical interface of p53 and is stabilized by hydrocarbon linkers (Figure 1.3C).[39] Foldamers are a chemically synthesized molecules that mimics the folded conformation of a biological macromolecule, but do not contain the typical linkages found in the natural molecule.[40, 41] This biologically orthogonal structure allows helical foldamers to evade peptidases or proteases in solution, thus overcoming the basic challenges of peptide therapies. Mixed α/β-peptides are a commonly used foldamer to resemble the helical motif. The Gellman group utilized this strategy to disrupt the PPI necessary for HIV-1 membrane fusion (gp41 6-helix bundle formation) by synthesizing an α/β-peptide that resembles the Cpeptide binding interface of gp41 (Figure 1.3D).[42, 43] Proteomimetics are a simplified helical interface mimetic that works by occupying the identical R group position of an α-helix without mimicking the entire helical structure. Hamilton et al. synthesized a 3,2',2'-functionalised terphenyl derivative to mimic the Cpeptide binding face of gp41 to inhibit HIV-1 membrane fusion.[44]

While these chemical methods for helical mimicking and stabilization have been successful as therapeutic designs over the last 15 to 20 years, they have significant drawbacks from a production and pharmacokinetic standpoint. Chemical synthesis is not only a tedious and costly method for therapeutic production, as the yield generally decreases with increasing molecular size, but because these synthetic routes produce smaller molecules (< 30 kDa) they are excreted through the renal clearance pathway rapidly, thus limiting the effectiveness of the therapies long term.[45]

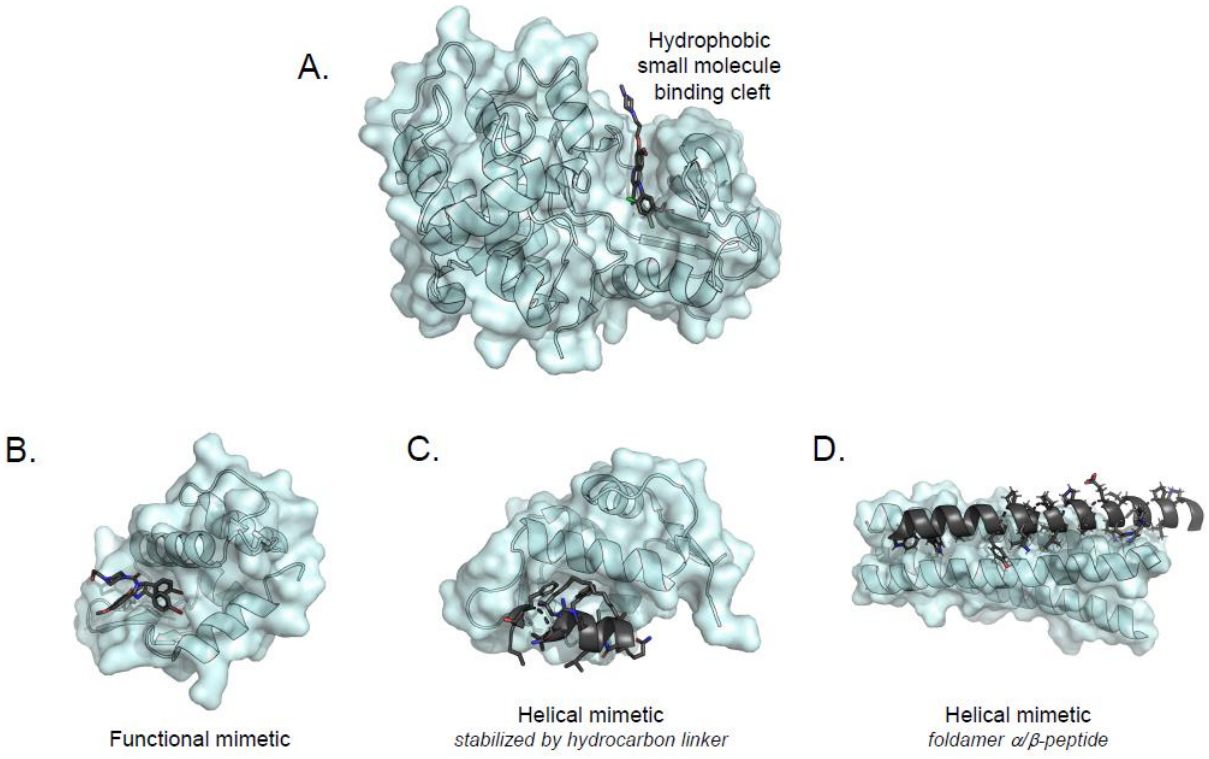


Figure 1.3 (A) Canonical small molecule in a well-defined hydrophobic binding pocket. (Example of human Src kinase bound to kinase inhibitor bosutinib, PDB: 4MXO) (B) Crystal structure of Nutlin-2 and MDM2 complex. Representation of a functional helix mimetic which occupies a helical binding site (PDB: 1RV1). (C) Crystal structure of the stapled p53 peptide bound to MDM2. Representation of a helical mimic stabilized by a hydrocarbon linker (PDB: 3V3B). (D) Crystal structure of a chimeric α + α/β-peptide analogue of the Cpeptide domain of gp41 in complex with the Npeptides of gp41. Representation of a helical mimetic stabilized by a foldamer backbone arrangement (PDB: 3G7A).

1.4 Protein Grafting for Helical Stabilization

An alternative helix stabilization method utilizes protein engineering to solubilize and stabilize a peptide mimic on a well-understood tertiary fold.[46] The design starts with a protein scaffold that contains a solvent-exposed α-helix. Based on the overall architecture of the scaffold protein, the peptide mimic could range from a small coiled-coil motif to a larger helical bundle consisting of three to four stabilizing helical assemblies.[47, 48] Some protein scaffolds contain a mixture of α-helices and beta sheets arranged in such a way that the solvent-exposed helical ligand

is cradled by the tertiary arrangement, thus enhancing the helical stability and promoting helical propensity.[49, 50] The benefit to this strategy is that the peptide mimic is easily expressible in *E. coli* (a common bioproduction host), is more resistant to proteolytic degradation and, depending on size, can improve pharmacokinetics *in vivo* compared to a peptide, peptide mimetic or small molecule.[45]

The Kim lab and Schepartz lab paved the way for protein grafts in developing unique techniques to display binding interfaces on small tertiary structures. This endows greater stability of peptides and resistance to proteolytic degradation. Schepartz lab designed 'miniature proteins' with a solvent-exposed α-helix supported by a single beta strand (Figure 1.4A).[49, 51] In grafting on a disease-relevant helical peptide, these structurally sound proteins can specifically target and inhibit disease-relevant PPIs. One such miniature protein, p007, utilized the helical binding interface of GCN4 grafted onto a small but stable avian pancreatic polypeptide (aPP) to generate a grafted protein with nanomolar binding affinity for DNA recognized by GCN4.[49] Kim et al. generated peptide dimers in a coiled-coil arrangement to utilize as a scaffold protein structure on which to graft the disease-relevant peptide (Figure 1.4B).[47] These dimers include solvent-exposed hydrophilic residues which increase solubility of the protein. The dimeric complex allows for greater stability of the protein. In one such study, Kim and colleagues grafted gp41 Cpeptide onto this coiled-coil to generate an antiviral inhibitor for HIV-1 membrane fusion. While these studies set the protein grafting technique in motion, they both utilize peptide synthesis to produce the therapeutics, thus limiting the cost-effective nature of the design.[47]

More recently affibodies emerged as a useful structural motif for engineering stabilized helical grafts. Affibodies are small, single domain proteins (58 amino acids, ~6 kDa) arising from the B-domain of Protein A and consisting of a three-helix bundle (Figure 1.4C).[48, 52] They have

been designed to maintain structural integrity in harsh environments and their small size has made them useful as imaging tools with chemical conjugation. Feldwisch et al. optimized affinity proteins using an affibody as a molecular scaffold. They adapted the solvent exposed residues on one helical ligand to resemble the peptide $Z_{HER2:342}$ which binds the human epidermal growth factor receptor 2 (HER2) involved in many cancers, including breast cancer.[53] While these scaffolds have proven enormously useful in generating high affinity and specificity in protein therapeutics, they are limited by the length of the peptide binding interface.

In another approach, our lab used Pleckstrin Homology (PH) domains as the foundation for the helical graft. PH domains have a solvent-exposed C-terminal α-helix cradled by tertiary beta sheets which allow it to be useful in grafting on important helical residues on the existing helix as well as extending the helix to cover a range of helical binding interfaces (Figure 1.4D).[50] In this protein engineering strategy, termed 'helix-grafted display' we focused on designing Cpeptide mimics that binds to and inhibits the gp41 membrane fusion protein in HIV-1 with an IC_{50} comparable to the FDA approved peptide inhibitor Enfuvirtide, but with improved *in vivo* stability.[54, 55]

Helix-grafting has proven to be a highly functional and cost-effective method for helix-stabilization. The peptide mimics generally show improved resistance to proteolytic degradation and similar binding affinities to the original peptide. Enhancing the binding profiles for these proteins is usually the next step from a protein engineering standpoint.

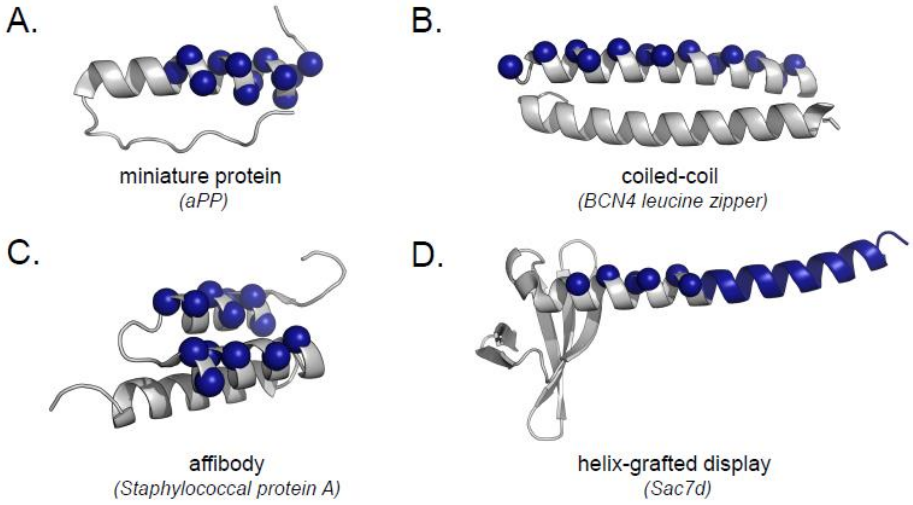


Figure 1.4 (A) Helix-grafting onto a miniature aPP protein. Density blue spheres indicate sites that have been mutated on the original scaffold (PDB: 2BF9). (B) Helix-grafting onto a coiled-coil leucine zipper motif. Density blue spheres indicate sites that have been mutated on the original scaffold (PDB: 4DMD). (C) Helix-grafting onto an affibody protein-A. Density blue spheres indicate sites that have been mutated on the original scaffold (PDB: 1Q2N). (D) Helix-grafting onto a PH-like Domain, Sac7d. Density blue spheres indicate sites that have been mutated on the original scaffold and density blue ribbon represents helical extension from the native helix (PDB: 2XIW and 1AIK).

1.5 Evolving Helical Proteins

Grafting existing ligands onto a protein is a good starting point in engineering novel PPIs or inhibitors, but the binding affinities for some of these interactions have room for improvement, as nature doesn't always select for the tightest affinities in a dynamic system. Evolving helical proteins for improved recognition of a PPI has been a major focus of recent research in probing and developing potential therapeutics. Selection and screening methods for proteins/helical ligands have varied and include tools like phage display, ribosome display, bacterial display, yeast display, yeast two-hybrid, and many others are still being developed. This review will cover some of the screening methods used to generate more potent helical grafts.

Phage display is still the most widely utilized method for screening a diverse protein library (Figure 1.5A).[56] The protein library is genetically fused to a phage coat protein and is therefore displayed on the surface of the assembled phage. The protein library can be screened for binding to a particular protein of interest by biopanning *in vitro* and amplifying enriched clones in *E. coli*.[56] In affibody design, phage display libraries were created and selected for by biopanning against a specific target or receptor.[46, 48, 53] This research has generated affibodies with affinities ranging from picomolar to micromolar binding interactions against targets such as insulin, serum albumin, HER2 receptor, just to name a few. The highest affinity-matured affibody protein to date is one that targets the Erb2 domain of HER2 with a K_D of 22 pM from Orlova et al.[53]

Ribosome display and bacterial display are less common but highly effective methods for improving binding profiles for helical proteins. Ribosome display allows the RNA encoding a specific library member to be tethered to the functional protein in which it encodes by stopping the translational process on a ribosome and screening for high affinity binders *in vitro*. The enriched clones can be reverse-transcribed and the DNA library amplified by a PCR reaction (Figure 1.5B).[57] Bacterial display utilizes the proteins on *E. coli* that are transported and exposed to the extracellular space. The protein library is genetically fused to this membrane protein and displayed on the outside of *E. coli*. Screening for high affinity binders to a specific protein of interest can be done using Fluorescence Activated Cell Sorting (FACS) (Figure 1.5C).[58, 59] Ankyrin repeat domains are a challenging scaffold to diversify with their repetitive nature. N- and C-terminal capping repeats were designed in work by Binz et al. to avoid aggregation, and these protein libraries were displayed and selected for in *E. coli* using ribosome display.[60] This yielded two kinase inhibitors (JNK2 and p38) with improved affinity in the nanomolar range.

Yeast display developed by the Wittrup lab has become an emerging tool for screening diverse protein libraries, as it utilizes a eukaryotic host with chaperones to improve not only affinity, but also to enrich for protein stability and expression (Figure 1.5D).[61, 62] In yeast display, much like bacterial display, the protein library is genetically fused to a membrane-associated protein in yeast and exported to the outside of the cell. Screening for high affinity binding events can be accomplished using FACS.[58] Our lab utilized yeast display to develop a polyclonal mixture of helix-grafted display proteins that inhibit HIV membrane fusion and potentially prevent resistance of the virus to the protein cocktail.[55, 63] The library generated peptide mimics with improved affinity (~2.6-6.2 nM) that retained HIV-1 inhibition across three strains of HIV-1 including Enfuvirtide-resistant strains.[55]

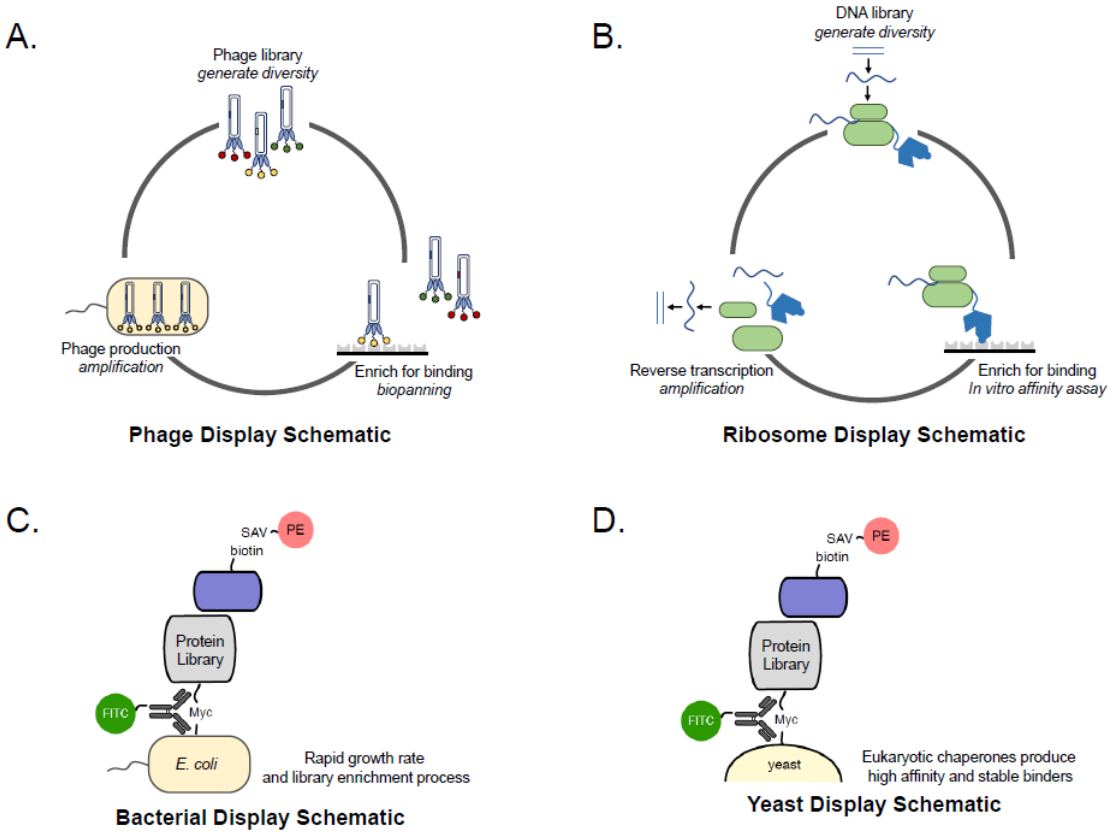


Figure 1.5 (A) Phage display schematic. Phage go through a cycle of creating diversity, biopanning for high affinity binders and amplification in *E. coli*. (B) Ribosome display schematic. DNA library is transcribed into RNA and translated on the ribosome where it remains tethered to the translated protein and enriched for high affinity binders before reverse-transcribing RNA into DNA and amplifying the enriched library. (C) Bacterial display schematic. Protein libraries are fused to a membrane-associated bacterial protein and displayed on the outside of bacteria where they can be screened using FACS for high affinity binders. (D) Yeast display schematic. Protein libraries are fused to a membrane-associated yeast protein (AGAII) and displayed on the outside of yeast cells where they can be screened using FACS for high affinity binders.

1.6 Conclusion

Helical ligands are ubiquitous in nature and have a profound impact on disease state of cellular function and signaling. Because of this, research has focused on probing and inhibiting these interactions involving an α-helix. The most common method for probing these PPIs is to mimic the native α-helix either through production of a helical peptide or peptide mimic. Many of

the mechanisms to develop a peptide mimic involve chemical synthesis which has the drawback of poor production yields and high production costs.

Overcoming this challenge led researchers to utilize protein fusions to solubilize peptides. This includes fusions of hyperstable proteins with protease cleavage sites to express the peptide in a common bioproduction host and then isolate the peptide post-purification. Fusions solve the high synthetic production costs, however peptides without a tertiary structure have a limited shelf life and poor pharmacokinetics due to their susceptibility to proteolytic degradation and rapid *in vivo* clearance. Protein fusions have extended into the realm of 'protein grafting' by taking a native protein containing a solvent-exposed helix and mutating the native helix to resemble the peptide of interest. Proteins used as scaffolds in protein grafting have ranged from miniature proteins and coiled-coils to larger tertiary structures like affibodies, ankyrin motifs and PH domains. Grafting on the native peptide-of interest aides in probing the helical PPI, but much of the time the native binding interaction has modest affinity for its receptor.

Evolving helical interfaces has bridged the gap from the modest affinity of the initial interaction to a much tighter binder, able to outcompete the wildtype peptide. Typically, phage display has been utilized for screening protein libraries. This has yielded improved affibodies and other helical bundles. Yeast display has also emerged as an impressive screening method as it contains chaperones to aide in protein folding and generally enriches for proteins with improved stability and expression as well as tighter affinity for the target. A variety of novel techniques for screening protein libraries are being utilized for protein evolution including ribosome display and bacterial display.

The combined progression in generating helical mimics that display improved binding affinity and *in vivo* stability has led to the research in this dissertation. This work takes the current

concepts of helical stabilization and improves upon the concepts, focusing on HIV-1 membrane fusion as a case study. As technology develops, targeting PPIs involving α-helices will be at the forefront of therapeutic development and protein engineering strategies.

REFERENCES

1. Pace, C.N. and J.M. Scholtz, *A helix propensity scale based on experimental studies of peptides and proteins.* Biophys J, 1998. **75**(1): p. 422-7.
2. Banga, A.K., *Therapeutic peptides and proteins : formulation, processing, and delivery systems*. Third edition. ed. 2015, Boca Raton, FL: CRC Press,Taylor & Francis Group. xvii, 374 pages, 8 unnumbered pages of plates.
3. Lynn, D.G. and S.C. Meredith, *Review: model peptides and the physicochemical approach to beta-amyloids.* J Struct Biol, 2000. **130**(2-3): p. 153-73.
4. Walsh, G., *Biopharmaceutical benchmarks 2014.* Nat Biotechnol, 2014. **32**(10): p. 992-1000.
5. Moasser, M.M., *Two dimensions in targeting HER2.* J Clin Oncol, 2014. **32**(19): p. 2074-7.
6. Deb, S.P., *Mutant P53 and MDM2 in cancer*. 2014, New York: Springer. pages cm.
7. Gadepalli, V.S., et al., *Lung cancer stem cells, p53 mutations and MDM2.* Subcell Biochem, 2014. **85**: p. 359-70.
8. Shangary, S. and S. Wang, *Targeting the MDM2-p53 interaction for cancer therapy.* Clin Cancer Res, 2008. **14**(17): p. 5318-24.
9. Chan, D.C., et al., *Core structure of gp41 from the HIV envelope glycoprotein.* Cell, 1997. **89**(2): p. 263-73.
10. Chan, D.C. and P.S. Kim, *HIV entry and its inhibition.* Cell, 1998. **93**(5): p. 681-4.

11. Jiao, J., et al., *Kinetically coupled folding of a single HIV-1 glycoprotein 41 complex in viral membrane fusion and inhibition.* Proc Natl Acad Sci U S A, 2015. **112**(22): p. E2855-64.

12. Care, A., P.L. Bergquist, and A. Sunna, *Solid-Binding Peptides in Biomedicine.* Adv Exp Med Biol, 2017. **1030**: p. 21-36.

13. Marston, F.A., *The purification of eukaryotic polypeptides synthesized in Escherichia coli.* Biochem J, 1986. **240**(1): p. 1-12.

14. Hannig, G. and S.C. Makrides, *Strategies for optimizing heterologous protein expression in Escherichia coli.* Trends Biotechnol, 1998. **16**(2): p. 54-60.

15. LaVallie, E.R. and J.M. McCoy, *Gene fusion expression systems in Escherichia coli.* Curr Opin Biotechnol, 1995. **6**(5): p. 501-6.

16. Terpe, K., *Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems.* Appl Microbiol Biotechnol, 2003. **60**(5): p. 523-33.

17. Raran-Kurussi, S. and D.S. Waugh, *Unrelated solubility-enhancing fusion partners MBP and NusA utilize a similar mode of action.* Biotechnol Bioeng, 2014. **111**(12): p. 2407-11.

18. Catanzariti, A.M., et al., *An efficient system for high-level expression and easy purification of authentic recombinant proteins.* Protein Sci, 2004. **13**(5): p. 1331-9.

19. LaVallie, E.R., et al., *A thioredoxin gene fusion expression system that circumvents inclusion body formation in the E. coli cytoplasm.* Biotechnology (N Y), 1993. **11**(2): p. 187-93.

20. Satakarni, M. and R. Curtis, *Production of recombinant peptides as fusions with SUMO.* Protein Expr Purif, 2011. **78**(2): p. 113-9.

21. Baker, R.T., *Protein expression using ubiquitin fusion and cleavage.* Current Opinion in Biotechnology, 1996. **7**(5): p. 541-546.

22. Butt, T.R., et al., *SUMO fusion technology for difficult-to-express proteins.* Protein Expression and Purification, 2005. **43**(1): p. 1-9.

23. Butt, T.R., et al., *Ubiquitin fusion augments the yield of cloned gene products in Escherichia coli.* Proceedings of the National Academy of Sciences of the United States of America, 1989. **86**(8): p. 2540.

24. Kim, K.I., S.H. Baek, and C.H. Chung, *Versatile protein tag, SUMO: Its enzymology and biological function*. 2002: New York. p. 257-268.

25. Moon, G.S., Y.R. Pyun, and W.J. Kim, *Expression and purification of a fusion-typed pediocin PA-1 in Escherichia coli and recovery of biologically active pediocin PA-1.* Int J Food Microbiol, 2006. **108**(1): p. 136-40.

26. Pryor, K.D. and B. Leiting, *High-level expression of soluble protein in Escherichia coli using a His6-tag and maltose-binding-protein double-affinity fusion system.* Protein Expr Purif, 1997. **10**(3): p. 309-19.

27. Smith, D.B. and K.S. Johnson, *Single-step purification of polypeptides expressed in Escherichia coli as fusions with glutathione S-transferase.* Gene, 1988. **67**(1): p. 31-40.

28. Uhlen, M., et al., *Gene fusion vectors based on the gene for staphylococcal protein A.* Gene, 1983. **23**(3): p. 369-78.

29. Clemens, A., et al., *Npro fusion technology to produce proteins with authentic N termini in E. coli.* Nature Methods, 2007. **4**(12): p. 1037.

30. Floss, D.M., et al., *Elastin-like polypeptides revolutionize recombinant protein expression and their biomedical application.* Trends in Biotechnology, 2010. **28**(1): p. 37-45.

31. Xiaopeng, Z., et al., *Recombinant Expression and Characterization of α-Conotoxin LvIA in Escherichia coli.* Marine Drugs, 2016. **14**(1): p. 11.

32. Zhao, Q., et al., *Recombinant production of medium- to large-sized peptides in Escherichia coli using a cleavable self-aggregating tag.* Microb Cell Fact, 2016. **15**(1): p. 136.

33. Vassilev, L., D. Fry, and SpringerLink, *Small-molecule inhibitors of protein-protein interactions*. 2011, Heidelberg

34. Azzarito, V., et al., *Inhibition of alpha-helix-mediated protein-protein interactions using designed molecules.* Nat Chem, 2013. **5**(3): p. 161-73.

35. Vassilev, L.T., et al., *In vivo activation of the p53 pathway by small-molecule antagonists of MDM2.* Science (New York, N.Y.), 2004. **303**(5659): p. 844.

36. Grasberger, B.L., et al., *Discovery and cocrystal structure of benzodiazepinedione HDM2 antagonists that activate p53 in cells.* Journal of medicinal chemistry, 2005. **48**(4): p. 909.

37. Sanjeev, S., et al., *Temporal activation of p53 by a specific MDM2 inhibitor is selectively toxic to tumors and leads to complete tumor growth inhibition.* Proceedings of the National Academy of Sciences, 2008. **105**(10): p. 3933.

38. Valeria, A., et al., *Inhibition of α-helix-mediated protein–protein interactions using designed molecules.* Nature Chemistry, 2013. **5**(3): p. 161.

39. Carvajal, L., et al., *Dual Inhibition of Mdmx and Mdm2 Using an Alpha-Helical P53 Stapled Peptide (ALRN-6924) As a Novel Therapeutic Strategy in Acute Myeloid Leukemia.* Blood, 2017. **130**(s1).

40. Tomasini, C., et al., *Foldamers.* European Journal of Organic Chemistry, 2013. **2013**(17): p. 3408-3409.

41. Gellman, S., *Foldamers: A manifesto*, in *Accounts Chem. Res.* 1998. p. 173-180.

42. Horne, W.S., et al., *Structural and biological mimicry of protein surface recognition by α/β-peptide foldamers.* Proceedings of the National Academy of Sciences, 2009. **106**(35): p. 14751.

43. Johnson, L.M., et al., *Enhancement of α-helix mimicry by an α/β-peptide foldamer via incorporation of a dense ionic side-chain array.* Journal of the American Chemical Society, 2012. **134**(17): p. 7317.

44. Orner, B.P., J.T. Ernst, and A.D. Hamilton, *Toward proteomimetics: terphenyl derivatives as structural and functional mimics of extended regions of an alpha-helix.* Journal of the American Chemical Society, 2001. **123**(22): p. 5382.

45. Fan, J. and I.A.M. de Lannoy, *Pharmacokinetics.* Biochemical Pharmacology, 2014. **87**(1): p. 93-120.

46. Hosse, R.J., A. Rothe, and B.E. Power, *A new generation of protein display scaffolds for molecular recognition*. 2006: Bristol. p. 14-27.

47. Samuel, K.S. and S.K. Peter, *Protein grafting of an HIV-1-inhibiting epitope.* Proceedings of the National Academy of Sciences of the United States of America, 2003. **100**(17): p. 9756.

48. Nygren, P.Å., *Alternative binding proteins: Affibody binding proteins developed from a small three-helix bundle scaffold*. 2008: Oxford, UK. p. 2668-2676.

49. Yang, L. and A. Schepartz, *Relationship between folding and function in a sequence-specific miniature DNA-binding protein.* Biochemistry, 2005. **44**(20): p. 7469.

50. Walker, S.N., et al., *GLUE That Sticks to HIV: A Helix-Grafted GLUE Protein That Selectively Binds the HIV gp41 N-Terminal Helical Region.* ChemBioChem, 2015. **16**(2): p. 219-222.

51. Kritzer, J.A., et al., *Miniature Protein Inhibitors of the p53–hDM2 Interaction.* ChemBioChem, 2006. **7**(1): p. 29-31.

52. Karin, N., et al., *Binding proteins selected from combinatorial libraries of an α-helical bacterial receptor domain.* Nature Biotechnology, 1996. **15**(8): p. 772.

53. Orlova, A., et al., *Tumor Imaging Using a Picomolar Affinity HER2 Binding Affibody Molecule.* Cancer Research, 2006. **66**(8): p. 4339-4348.

54. Tennyson, R.L., et al., *Helix-Grafted Pleckstrin Homology Domains Suppress HIV-1 Infection of CD4-Positive Cells.* ChemBioChem, 2016. **17**(20): p. 1945-1950.

55. Ikeda, T., Tennyson, R.L., Walker, S.N., Harris, R.S., McNaughton, B.R., *Evolved Biologics Designed for Monthly Administration Potently Inhibit Entry of Enfuvirtide-Resistant HIV-1.* ACS Infectious Disease, 2019.

56. Shim, H., *Antibody Phage Display.* Advances in experimental medicine and biology, 2017. **1053**: p. 21.

57. Jozef, H. and P. Andreas, *In vitro selection and evolution of functional proteins by using ribosome display.* Proceedings of the National Academy of Sciences of the United States of America, 1997. **94**(10): p. 4937.

58. Stowe, C., et al., *Flow-Based Single Cell Deposition for High-Throughput Screening of Protein Libraries.* PLoS One, 2015. **10**(11): p. e0140730.

59. Kenrick, S.A. and P.S. Daugherty, *Bacterial display enables efficient and quantitative peptide affinity maturation.* Protein Engineering, Design & Selection, 2010. **23**(1): p. 9-17.

60. Forrer, P., et al., *A novel strategy to design binding molecules harnessing the modular nature of repeat proteins.* FEBS Letters, 2003. **539**(1-3): p. 2-6.

61. Liu, B. and SpringerLink, *Yeast surface display : methods, protocols, and applications*. 2015, New York: New York : Humana Press.

62. Gai, S.A. and K.D. Wittrup, *Yeast surface display for protein engineering and characterization.* Current Opinion in Structural Biology, 2007. **17**(4): p. 467-473.

63. Tennyson, R.L., et al., *Evaluation of sequence variability in HIV-1 gp41 C-peptide helix-grafted proteins.* Bioorganic & Medicinal Chemistry, 2018. **26**(6): p. 1220-1224.

CHAPTER TWO

Helix-grafted Display: HIV-1 as a Case Study for Helical Stabilization Using a Protein Engineering Strategy[1]

In this work, the protein engineering design and initial analysis was done by myself. Rachel Tennyson aided me in cloning and protein purification as well as protein expression and binding analysis experiments. I also carried out serum stability assays and ELISA experiments.

2.1 Introduction

The ongoing discovery of new drug therapies is of vital importance to human health. The traditional pharmaceutical paradigm for this discovery centers on small molecules binding to well defined protein pockets, typified by enzyme active sites. However, there remain countless important targets largely beyond the reach of this strategy, principally due to extended contact surfaces. Such interactions are often collected under the heading of "protein–protein interactions" or PPIs.[1, 2]

A medicinally significant subset of these PPIs feature binding of one protein to an exposed helix on another, which has sparked considerable interest in the synthetic replication of helical epitopes as a route to novel therapeutics. Various strategies have been employed, including oligomeric organic scaffolds that project side chains along appropriate vectors,[3-5] covalently constrained (or "stapled") peptides,[6-9] and helical "foldamers"[10-14] employing natural or

[1] Reprinted with permission from:
Walker, S.N., Tennyson, R.T., Chapman, A.M., Kennan, A.J., McNaughton, B.R., GLUE That Sticks to HIV: A Helix-Grafted GLUE Protein That Selectively Binds the HIV gp41 N-Terminal Helical Region, *ChemBiochem*, 2015, 16, 219

unnatural backbone architectures. Each has produced effective agents, but all require nontrivial synthetic effort and expense.

A more accessible suite of ligands might be had by exploiting the flexibility and feasibility of protein expression. If one could identify a scaffold protein bearing a helix with at least one solvent-exposed face, and if that protein was simple and stable enough to facilitate easy expression and tolerate varying the exposed helical residues and helix length, it could serve as a generic scaffold upon which to graft any desired helical interface. The result would be a protein graft in which residues critical to recognition of a particular epitope are mutated onto the host protein in appropriate positions along the solvent-exposed helix. The general concept of grafting is an established method for mimicking protein surfaces.[15-19]

In this proof-of-concept manuscript, we report successful application of our specific "helix-grafting" technique to a PPI crucial for HIV infection. The HIV fusion protein gp41 operates in part by binding a C-terminal helix (Cpeptide) onto a trimeric N-terminal helix in a coiled-coil arrangement. We show that grafting of gp41 Cpeptide residues onto the exposed helical face of a suitable protein scaffold affords a new ligand that expresses well in *E. coli*, exhibits excellent serum stability, and is capable of replicating the native interaction. Following initial attachment to a target cell, HIV entry is affected by fusion of the respective cell membranes, mediated by gp41.[20, 21] In the prefusogenic state, gp41 is trimerized, and an N-terminal fusion peptide from each subunit inserts into the target membrane. The protein then undergoes a conformational rearrangement in which the Cpeptide packs as an antiparallel helix against the surface of the N-terminal Heptad Repeat (NHR) trimer.[22, 23] Anchoring of the respective protein termini into viral and cell membranes ensures that this rearrangement requires membrane juxtaposition and thus provides a mechanism for promoting fusion (Figure 2.1A).[24-27] Peptides derived from

either the Npeptide helix or Cpeptide helix (some as short as 12–16 residues) have been shown to bind the coiled-coil and inhibit membrane fusion by HIV.[24-31] The best known of these, enfuvirtide (marketed as Fuzeon™), consists of 36 residues and is an FDA-approved treatment (Figure 2.1B).[32, 33] However, like other short peptide drugs, its chemical synthesis is extraordinarily expensive and it exhibits poor serum stability ($t_{1/2}$ ~3.8 h).[34] We reasoned that a helix-grafted alternative might retain similar specificity but have improved stability, solubility, and availability.

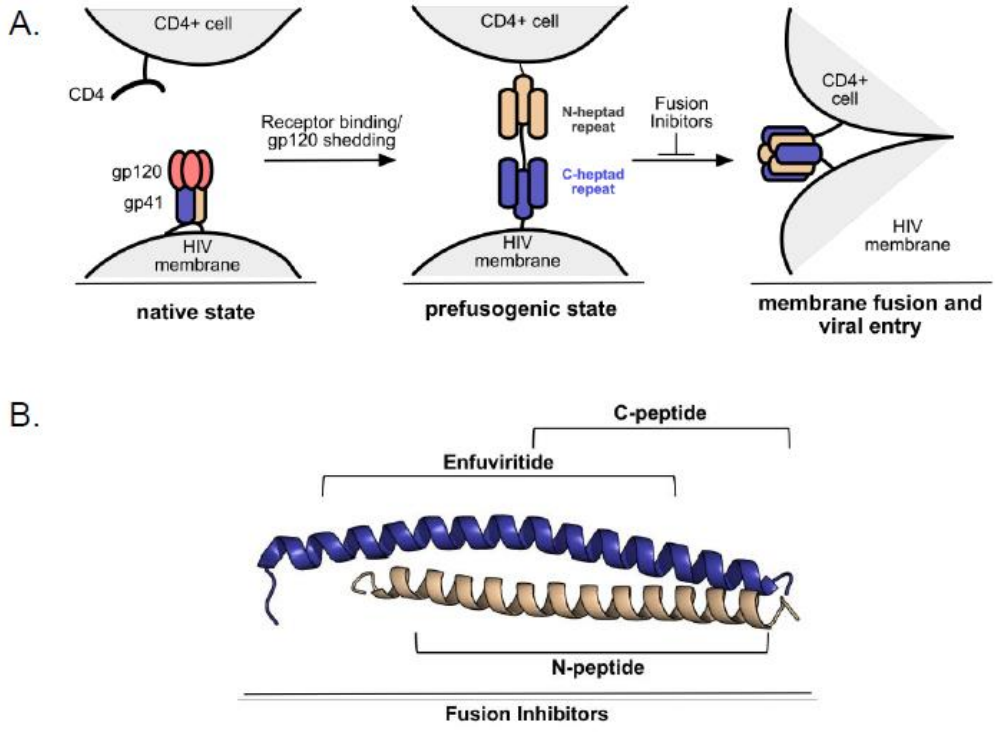


Figure 2.1 (A) HIV viral fusion by formation of a prefusogenic state leading to a trimer-of-hairpins assembly involving the N-terminal helical region (NHR, tan) and the C-terminal helical region (CHR, density blue) of gp41. Fusion inhibitors bind to the prefusogenic state and prevent the formation of the trimer-of-hairpins assembly. (B) Crystal structure of the NHR/CHR monomeric complex (PDB ID: 1AIK). Sequences used as fusion inhibitors are denoted as C-peptide, Enfuvirtide and N-peptide.

2.2 Helix-grafted Display of Cpeptide on GRAM-Like Ubiquitin-binding in EAP45 (GLUE)

In designing our first-generation helix-grafted protein, we focused on a Pleckstrin homology (PH) domain called GLUE (GRAM-Like Ubiquitin-binding in EAP45, Figure 2.2A, gray), which is derived from a subunit of the endosomal sorting complex.[35] Like other members of the PH family, the GLUE domain contains a C-terminal amphipathic helix resting in a cleft formed by two opposing β-sheets, with one face presented to solvent. This relatively rare arrangement is well suited to serve as a helix-grafting scaffold. Although the native GLUE helix is only 16 residues, known structures of other PH domains with helices up to 29 residues suggested that it could be extended to a length comparable to the Cpeptide without structural compromise. In addition to its well-positioned helix platform, GLUE is a relatively small (~15 kDa) and stable protein. Finally, native GLUE function relies on an affinity for phospholipids that can be abolished by a single Arg107Ala mutation, making it suitable for future intracellular targets without fear of disrupting lipid trafficking.[35]

We began by aligning the native helix on GLUE (Figure 2.2A, grey) with a single C-peptide helix from gp41 (Figure 2.2A, density blue). Backbone atoms from the GLUE helix (PDB ID: 2CAY) were aligned with the corresponding number from the N-terminal segment of the gp41 C-peptide (PDB ID: 1AIK) by using the PyMOL pair_fit algorithm. The overlay was very good (RMSD of 0.44 over 60 atoms), and it allowed us to confidently select six positions on the GLUE helix at which we could install side chains from the gp41 sequence in such a way as to replicate their native three-dimensional positions. We then extended the helix by genetically extending a pure gp41 sequence to the C-terminus of GLUE (Figure 2.2A), such that the total length of the new helix was appropriate for binding to the trimeric N-terminal coiled-coil. The helix-grafted protein showed some steric hindrance from phenylalanine 24 (F24) on the beta sheet portion with

a tryptophan on the grafted helix. The phenylalanine was mutated to an alanine (F24A) to prevent any folding constraints from the van der waals forces (Figure 2.2B, red spheres). The final sequence was expressed as a soluble protein in *E. coli* (Figure 2.2C).

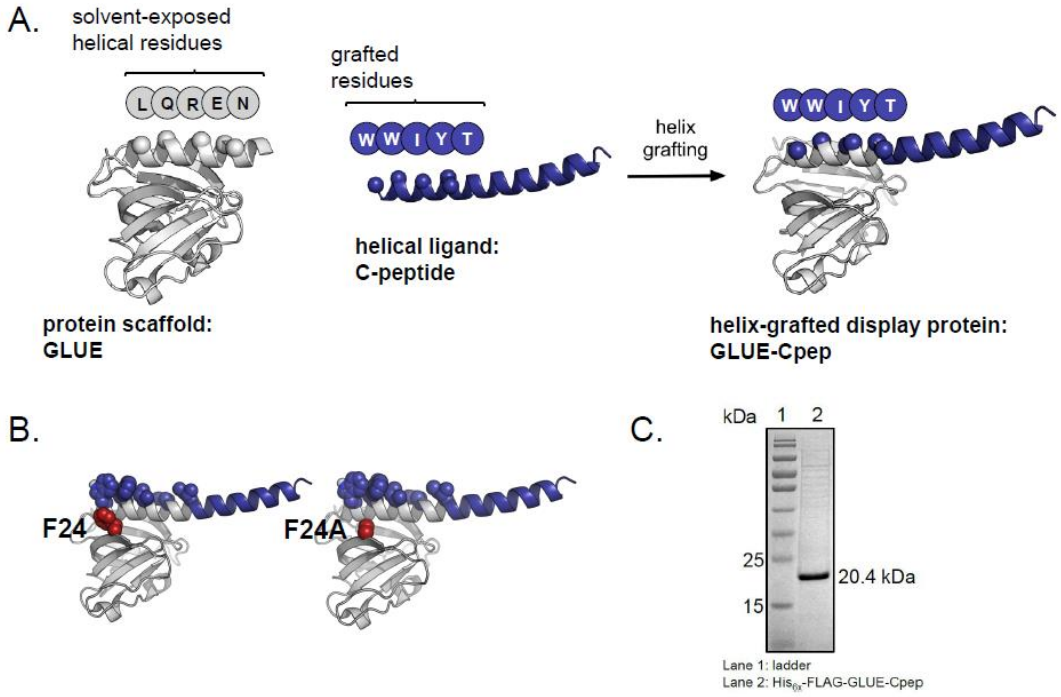


Figure 2.2 (A) Wild-type GLUE (gray) and gp41 C-peptide (dark blue). Sequences of the GLUE helix and the corresponding region of the gp41 helix are shown in gray and blue, respectively. Spheres indicate the Cα positions for each. Helix-grafted GLUE-Cpep, produced by backbone alignment of the independent structures. Spheres indicate Cα positions of GLUE residues mutated to those from gp41 (also color coded in the sequence). (B) Steric hindrance from residue F24 mutated to F24A to relieve folding strain. (C) Helix-grafted GLUE-Cpep (with F24A mutation) expressed as a soluble protein in *E. coli*.

2.3 Structural Analysis of Helix-grafted Display for Developing Stabilized Helical Ligands

We characterized both wild-type GLUE (wtGLUE) and the helix-grafted variant (referred to as GLUE-Cpep herein) by circular dichroism (CD) to probe for macroscopic structural changes (Figure 2.3A). Both proteins display a similar overall signal, suggesting that the grafting process does not compromise the GLUE domain fold. As one element of our design was the expectation that a well-folded protein domain would exhibit improved serum stability compared to an isolated

short peptide, we next conducted a serum stability test using a standard assay.[36] FLAG-tagged GLUE-Cpep, incubated with human serum for up to 12 h, showed no appreciable degradation by western blot analysis (Figure 2.3B). This supports a significant serum stability enhancement for the grafted protein compared to isolated peptides such as enfuvirtide.

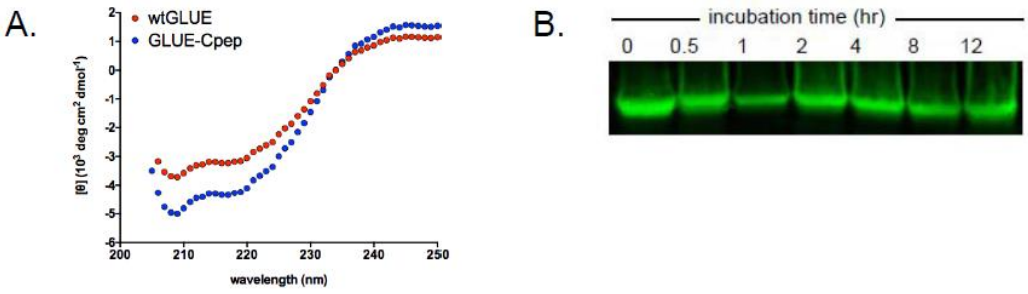


Figure 2.3 (A) Circular dichroism spectra for wildtype GLUE (•) and GLUE-Cpep (•). Wave traces show similarities in secondary structure with increased helicity for the helix-grafted GLUE-Cpep. (B) Western blot of FLAG-GLUE-Cpep prior to incubation with human serum (lane 1) and throughout incubation with human serum for 0.5 to 12 h (lane 2-7).

2.4 Functional Analysis of Helix-grafted Display for Developing Stabilized Helical Ligands

Direct analysis of the binding interaction between GLUE-Cpep and the NHR receptor by simple mixing of the two components is complicated by several factors: proper self-assembly of the N-terminal peptide, the potential for one, two, or three GLUE-derived ligands per complex, and the known susceptibility of unbound N-peptide trimers to aggregation/precipitation. Fortunately, these challenges have long been recognized, and several solutions exist. We chose to use a construct called 5-helix, based on initial work by Kim and co-workers.[37] It solves the problem of multiple equilibria and binding sites by covalently tethering five of the six subunits with short Gly/Ser loops. Thus, a single polypeptide contains three copies of the NHR domain and two C-peptides, such that when folded, it features the coiled-coil with two of its binding sites

already occupied and just a single exposed interface (Figure 2.4A). Throughout, we used 5-helix as a receptor to assess complex formation with GLUE-Cpep. Initial CD characterization of the GLUE-Cpep/5-helix complex demonstrates binding-induced gains in helicity and thermal stability. The wavelength spectrum (Figure 2.4B) exhibits a notably deeper signal for the 1:1 mixture than for either component alone, and the corresponding melt data (Figure 2.4C) reveals a dramatic increase in thermal stability, as evidenced by a significant shift of the overall melting curve, though the change in Tm is more modest (observed Tm values of ~77, ~79, and ~83°C for 5-helix, GLUE-Cpep, and the complex, respectively). The melting transition for the 1:1 sample is also highly cooperative, further supporting a well-defined assembly.

Having validated the GLUE-Cpep/5-helix interaction, we moved on to probe its viability in more complex environments. Binding in living cells (*E. coli*) was first assessed by split-superpositive green fluorescent protein (split-spGFP) reassembly, a technique we recently reported.[38, 39] *E. coli* were co-transformed with plasmids encoding 5-helix fused to the N-terminal half of spGFP (N-spGFP-5-helix) and one of two C-spGFP fusions: GLUE-Cpep or the gp41 C-peptide by itself. Interaction-dependent reassembly of GFP fragments (to generate a fluorescent signal) was measured by flow cytometry. Cells expressing either ligand construct are highly fluorescent, in contrast to a control with nothing fused to C-spGFP (Figure 2.4D). We further characterized this interaction by using ELISA. The grafted GLUE (GLUE-Cpep) bound 5-helix with slightly better affinity than the native C-peptide, whereas the wild-type GLUE exhibits no appreciable affinity thus confirming the need for the grafted domain (Figure 2.4E). This ELISA signal, indicating binding to a biotinylated 5-helix (Supplemental Figure 6.1A), was observed even for a GLUE-Cpep sample that was pre-incubated with human serum for up to 12 h (Figure 2.4F), confirming that the degradation-resistant form of the protein remains functional. Together, these

experiments show that the helix-grafted GLUE binds 5-helix in the context of a complex cellular milieu, in a manner comparable to the native ligand and with improved serum longevity compared to FDA-approved Enfuvirtide (~3.8 h).

Binding selectivity was assessed by measuring the amount of protein that was copurified from *E. coli* expressing an untagged GLUE-Cpep (~17.1 kDa) and His6-tagged 5-helix (~25.4 kDa). The tagged 5-helix copurified with a single protein, which was identified as GLUE-Cpep by mass spectrometry (Supplemental Figures 6.1B and 6.1C). The similar amounts of each copurified protein, as determined by densitometry measurements of each protein band, further indicated that the complex involved a 1:1 ratio of proteins (data not shown). The relatively miniscule levels of other copurified cellular proteins indicates excellent selectivity for this interaction, even in a complex cellular environment, suggesting a reasonably strong mutual affinity.[40]

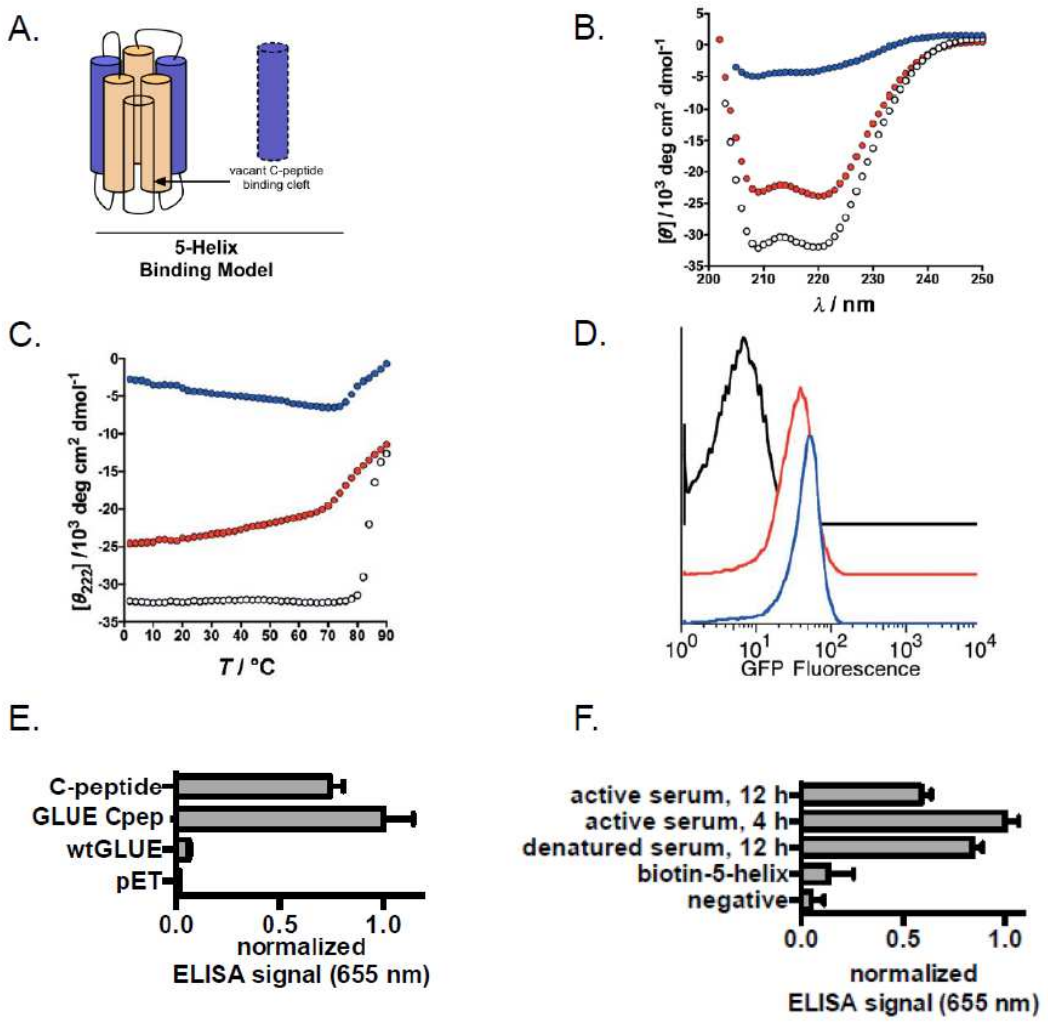


Figure 2.4 (A) Depiction of 5-helix, a single protein consisting of three copies of gp41 N-peptide (beige) and two copies of gp41 C-peptide (purple). When folded, this protein presents a single binding cleft (translucent column with dashed border) for a Cpeptide or Cpeptide mimic (purple column with dashed border). (B) Circular dichroism wave spectra of GLUE-Cpep (•), 5-helix (•), and a premixed 1:1 ratio of 5-helix and GLUE-Cpep (◦) (C) Circular dichroism thermal melt data at 222 nm for GLUE-Cpep (•), 5-helix (•), and a premixed 1:1 ratio of 5-helix and GLUE-Cpep (◦) showing temperature-dependent unfolding with increased stability for the complex (Tm ~79, ~77, and ~83°C for GLUE-Cpep, 5helix and the 1:1 complex). (D) Flow cytometry histograms for *E. coli* following split-spGFP reassembly experiments. (**-**) N-spGFP-5-helix/C-spGFP, (**-**) N-spGFP-5-helix/C-spGFP-Cpep, (**-**) N-spGFP-5-helix/C-spGFP-GLUE-Cpep. (E) *in cellulo* ELISA signal from *E. coli* cell lysate that expressed empty pETDuet plasmid, or pETDUET encoding for His6-5-helix and FLAG-wtGLUE, GLUE-Cpep, or Cpeptide. wtGLUE exhibits no appreciable signal while Cpeptide and GLUE-Cpeptide display appreciable binding. (F) in vitro ELISA on streptavidin (SA) coated plates with immobilized biotin-5-helix and FLAG-GLUE-Cpep incubated with denatured or active human serum. Appreciable binding is observed even after active serum incubation for up to 12 h.

2.5 Conclusion

In conclusion, we have demonstrated that the solvent exposed C-terminal α-helix of the GLUE protein scaffold can be dramatically modified and extended, so as to mimic the function of the gp41 C-peptide. ELISA and copurification data indicate that GLUE-Cpep selectively binds 5-helix, a protein that mimics the native C-peptide receptor. Unlike the isolated C-helix of enfuvirtide, GLUE-Cpep is soluble and well-folded in aqueous solution at room temperature (~25 ˚C) and is resistant to degradation in human serum at physiological temperature (~37 ˚C). Thus, this protein drug lead overcomes challenges faced by traditional peptide reagents and could represent a new reagent for inhibition of HIV entry. Additionally, helix-grafting onto PH and PH-like domains, such as GLUE, might be a general approach to the development of new reagents of interest to a diverse set of diseases that rely on helix-driven assembly. Finally, GLUE-Cpep serves as a starting point for the generation of higher affinity and more selective mutants through the application of high-throughput screening or selection methods. Such experiments are currently underway and will be reported in due course.

2.6 Methods

Protein Purification

Genes were cloned into pETDuet-1 using restriction enzymes BamHI and PacI, downstream of a His6x tag and transformed into BL21s (DE3). Cells were grown in 2.5 L LB cultures containing 50 μg/mL carbenicillin at 37 °C to OD_{600}=0.5 and induced with 1 mM IPTG at 25 °C overnight. Cells were then collected by centrifugation, resuspended in Tris buffer (20 mM Tris pH 7.4, 100 mM NaCl, 10 mM $(NH_4)_2SO_4$) and stored at -20 °C. Frozen pellets were thawed and sonicated with 1 second pulses for 2 minutes. The lysate was cleared by centrifugation (15,000 rpm 30 min.)

and the supernatant was mixed with 1 mL of Ni-NTA agarose resin for 1 hour at 4 °C. The resin was collected by centrifugation (4,750 rpm, 10 min.). The resin was sequentially washed with 50 mL of buffer containing 20 mM imidazole, 10 mL buffer containing 50 mM imidazole, and 5 mL buffer containing 100 mM imidazole. The protein was then eluted with 4 mL buffer containing 400 mM imidazole. The proteins were dialyzed against Tris buffer and analyzed for purity by SDS-PAGE. Purified protein concentrations were quantified using Beer's Law at an absorbance of 280 nm, following standard practice. In general, GLUE-Cpep was expressed as a soluble protein (~3 mg/L of *E. coli* culture)

Resolubilization from Inclusion Bodies

5Helix-His was cloned into a modified pETDuet-1 vector using restriction enzymes NdeI and KpnI and transformed into BL21s (DE3). Cells were induced to express 5Helix-His6x and lysed as described above. The lysate was cleared by centrifugation (15,000 rpm, 30 min.) and the supernatant discarded. The pellet was washed twice with Tris buffer containing 0.5 % Triton® X-100 and once with Tris buffer. The pellet was resuspended in urea buffer (Tris buffer with 8 M urea and 10 mM imidazole) to resolubilize the inclusion bodies and cleared by centrifugation (9,500 rpm, 30 min.) The supernatant was mixed with 1 mL of Ni-NTA agarose resin for 1 hour at 4 °C. The resin was collected by centrifugation (4,750 rpm, 4 min.). The resin was washed with 50 mL of urea buffer and eluted with 40 mL of urea elution buffer (Tris buffer with 6 M urea and 100 mM imidazole) into 500 mL Tris buffer by gravity dripping while stirring to refold the protein. The 540 mL elution was run through a column containing 1 mL of Ni-NTA agarose resin and eluted with 5 mL Tris buffer containing 400 mM imidazole. The protein was then eluted with 4 mL buffer containing 400 mM imidazole. The proteins were dialyzed against buffer and analyzed

for purity by SDS-PAGE and analyzed for refolding by CD. Purified proteins were quantified using Beer's Law at an absorbance of 280 nm.

Circular Dichroism

Proteins were purified as described above. Separately, each protein was diluted to 7-9 μM in Tris buffer. Wavelength data are the average of three scans from 250 nm to 200 nm in 1 nm steps at 25 °C. Thermal denaturation experiments at 222 nm were run from 0 to 90 °C in two-degree steps at a two-degree/minute rate of increase with one-minute equilibration and data averaging at each temperature. Tm values were obtained from minima of the first derivative of θ versus 1/T plots.

Serum Stability Assay

Using a previously described assay for serum stability, GLUE-Cpep was cloned into pET-28a(+) vector with an N-terminal FLAG tag (DDDDK) using restriction enzymes NdeI and HindIII. The completed construct was transformed into BL21s (DE3) and purified as described previously. 1 mL of RPMI supplemented with 25 % (v/v) of human serum was equilibrated at 37 °C. GLUE-Cpep was added to the solution to obtain a final concentration of 50 μg/mL and incubated at 37 °C. At known time intervals of 0.5, 1, 2, 4, 8, or 12 hours, 100 μL of the reaction solution was removed and denatured at 94 °C for 20 minutes and stored at -80 °C. Samples were separated by SDS-PAGE and transferred to a nitrocellulose membrane via an iBlot western blotting apparatus. The membrane was washed with PBS and incubated in LI-COR Blocking Buffer at 4 °C for 1 hour. The membrane was then incubated with a mouse anti-DDDDK antibody (anti-FLAG) in LI-COR Blocking Buffer for 1 hour at 4 °C. The membrane was washed 3x with PBS containing 0.1 % Tween-20, and then incubated with a IRDye 800CW Goat anti-mouse IgG-LI-COR secondary antibody in LI-COR Blocking Buffer for 1 hour at room temperature. The membrane was washed 3x with PBS containing 0.1 % Tween-20 and imaged using the Odyssey Classic Infrared Imager.

Split-Superpositive GFP (split-spGFP) Reassembly Assay

N-terminal superpositive GFP genetically tethered to the 5helix was cloned into pETDuet-1 using restriction enzymes BamHI and PacI. The completed construct was transformed into BL21s (DE3) and the cells were made electrocompetent via standard protocols. Separately, GLUE-Cpep and C-peptide tethered to the C-terminal fragment of superpositive GFP were independently cloned into pBAD vector using restriction enzymes NcoI and BsrGI. Constructs were electroporated into electrocompetent BL21s (DE3) containing the N-terminal superpositive GFP plasmid, pulsing at 1.8 kV in a 1 mm cuvette. Cells were allowed to recover in SOC media at 37 °C for 1 hour, and then plated onto agar plates containing carbenicillin and kanamycin. Individual colonies were picked and passaged once, followed by induction at 37 °C with 1 μM IPTG and 0.2 % arabinose when cultures reached an OD_{600} of 0.5. After 6 hours, cells were spun down and resuspended in 5 mL PBS. GFP fluorescence was measured by MoFlo Flow Cytometer.

***In cellulo* ELISA**

Separately, wtGLUE, GLUE-Cpep, and the Cpeptide were cloned into MCS1 of pETDuet-1 with FLAG tags using restriction enzymes NcoI and NotI. The 5helix with a C-terminal His6x tag was cloned into MCS2 of pETDuet-1 using restriction enzymes NdeI and KpnI. Completed constructs were transformed into BL21s (DE3). Cells containing the co-expressed pair were inoculated and induced as described previously. Cells were spun down and resuspended in lysis buffer (20 mM Hepes pH 7.5, 100 mM NaCl), lysed by sonication, and spun down to remove cell debris. Cleared lysates were incubated on clear Ni-NTA coated plates for 1 hour at room temperature and washed 4x with 200 μL wash buffer (20 mM Hepes pH 7.5, 150 mM NaCl, 0.05 % Tween-20, 0.01 mg/mL BSA). HRP-conjugated mouse anti-DDDDK antibody in LiCor Blocking Buffer was incubated

for 1 hour at room temperature, followed by 4x 200 µL washes. Colorimetry was developed using TMB-One substrate and absorbance was measured at 655 nm on a SynergyMx Microplate Reader.

ELISA Binding Assay in Human Serum

GLUE-Cpep was cloned into pET-28a(+) with an N-terminal FLAG tag using restriction enzymes NdeI and HindIII. 5helix was cloned into pETDuet-1 with an N-terminal AviTag and a C-terminal His6x tag using restriction enzymes NcoI and PacI. The completed constructs were transformed into BL21s (DE3) and purified as described previously. AviTag-5helix was conjugated to biotin using Avidity BioMix protocols and purified BirA Protein Ligase at 1.0 mg/mL. Biotin conjugation was confirmed by Mass Spectrometry. Separately, 5 mL of RPMI supplemented with 25 % (v/v) human serum, 5 mL of boiled RPMI supplemented with 25 % (v/v) of human serum, and 5 mL of (1x) PBS were equilibrated at 37 °C. GLUE-Cpep was added to each solution to a final concentration of 50 nM and incubated at 37 °C for 4 or 12 hours. 200 µL wash buffer (1x PBS pH 7.4, 0.1 % Tween-20, 0.02 mg/mL BSA) was incubated on clear streptavidin coated plates for 1 hour at room temperature to block. 100 µL of Biotinylated 5-helix at a concentration of 10 µg/mL was incubated for 1 hour at room temperature and washed 4x with 200 µL wash buffer. 100 µL of human serum-incubated GLUE-Cpep solutions were incubated on the plates for 1 hour at room temperature and washed 4x with 200 µL wash buffer. HRP-conjugated mouse anti-DDDDK antibody in LI-COR blocking buffer was incubated for 1 hour at room temperature, followed by 5x 200 µL washes. Colorimetry was developed using TMB-One substrate and absorbance was measured at 655 nm on a SynergyMx Microplate Reader.

Lysate Ni-NTA Pulldown Assay

GLUE-Cpep was cloned into MCS1 of pETDuet-1 using restriction enzymes NcoI and NotI. The 5helix with a C-terminal His6x tag was cloned into MCS2 of pETDuet-1 using the restriction

enzymes NdeI and KpnI. Completed constructs were transformed into BL21s (DE3). Cells containing the co-expressed pair were inoculated and induced as described previously. Cells were spun down and resuspended in lysis buffer (100 mM NaCl, 20 mM Tris pH 7.4, 10 mM $(NH_4)_2SO_4$), lysed by sonication, and spun down to remove cell debris. Cleared lysate was incubated with 200 µL Ni-NTA agarose resin for 1 hour. Ni-NTA agarose was washed with 8 mL lysis buffer containing 10 mM imidazole. Proteins were eluted with 500 µL lysis buffer containing 400 mM imidazole. The pulldown was analyzed by SDS-PAGE and confirmed by Mass Spectrometry.

2.7 Sequences of Proteins Used in this Work

For circular dichroism experiments

His6x-wtGLUE

MGSSHHHHHHSQDPEYWHYVETTSSGQPLLREGEKDIFIDQSVGLYHGKSKILQRQRGR
IFLTSQRIIYIDDAKPTQNSLGLELDDLAYVNYSSGFLTRSPALILFFKDPSSSTEFVQLSFR
KSDGVLFSQATERALENILT

5-Helix-His6x

MQLLSGIVQQQNNLLRAIEAQQHLLQLTVWGIKQLQARILAGGSGGHTTWMEWDREIN
NYTSLIHSLIEESQNQQEKNEQELLEGSSGGQLLSGIVQQQNNLLRAIEAQQHLLQLTVW
GIKQLQARILAGGSGGHTTWMEWDREINNYTSLIHSLIEESQNQQEKNEQELLEGSSGGQ
LLSGIVQQQNNLLRAIEAQQHLLQLTVWGIKQLQARILAGGHHHHHH

His6x-GLUE-Cpep

MGSSHHHHHHSQDPGLNDIFEAQKIEWHEGGSGGSGGTEYWHYVETTSSGQPLLREGE
KDIAIDQSVGLYHGKSKILQRQRGRIFLTSQRIIYIDDAKPTQNSLGLELDDLAYVNYSSG

FLTRSPALILFFKDPSSSTEFVQLSFRKSDGVWFSWATEIALYTILIHSLIEESQNQQEKNE

QELL

For ELISA experiments

FLAG-GLUE

MDYKDDDDKGGSEYWHYVETTSSGQPLLREGEKDIFIDQSVGLYHGKSKILQRQRGRIF

LTSQRIIYIDDAKPTQNSLGLELDDLAYVNYSSGFLTRSPALILFFKDPSSSTEFVQLSFRK

SDGVLFSQATERALENILT

FLAG-GLUE-Cpep

MDYKDDDDKGGSEYWHYVETTSSGQPLLREGEKDIAIDQSVGLYHGKSKILQRQRGRIF

LTSQRIIYIDDAKPTQNSLGLELDDLAYVNYSSGFLTRSPALILFFKDPSSSTEFVQLSFRK

SDGVWFSWATEIALYTILIHSLIEESQNQQEKNEQELL

FLAG-CHR

MDYKDDDDKGGSWMEWDREINNYTSLIHSLIEESQNQQEKNEQELL

5 Helix-His6x

(same as that used in Circular Dichroism)

For serum stability experiments

FLAG-GLUE-Cpep

MGSSHHHHHHSSGLVPRGSHMDYKDDDDKGGSEYWHYVETTSSGQPLLREGEKDIAID

QSVGLYHGKSKILQRQRGRIFLTSQRIIYIDDAKPTQNSLGLELDDLAYVNYSSGFLTRSP

ALILFFKDPSSSTEFVQLSFRKSDGVWFSWATEIALYTILIHSLIEESQNQQEKNEQELL

For split-superpositive GFP reassembly assay experiments

GLUE-Cpep-CscGFP

MEYWHYVETTSSGQPLLREGEKDIAIDQSVGLYHGKSKILQRQRGRIFLTSQRIIYIDDAK
PTQNSLGLELDDLAYVNYSSGFLTRSPALILFFKDPSSSTEFVQLSFRKSDGVWFSWATEI
ALYTILIHSLIEESQNQQEKNEQELLGGSGGSGTSGGSGKNGIKAKFKIRHNVKDGSVQL
ADHYQQNTPIGRGPVLLPRNHYLSTRSKLSKDPKEKRDHMVLLEFVTAAGIKHGRDER
YK

Cpeptide-CscGFP

MGWMEWDREINNYTSLIHSLIEESQNQQEKNEQELLGGSGGSGTSGGSGKNGIKAKFKI
RHNVKDGSVQLADHYQQNTPIGRGPVLLPRNHYLSTRSKLSKDPKEKRDHMVLLEFVT
AAGIKHGRDERYK

NscGFP-5helix

MGHHHHHHGGASKGERLFRGKVPILVELKGDVNGHKFSVRGEGKGDATRGKLTLKFIC
TTGKLPVPWPTLVTTLTYGVQCFSRYPKHMKRHDFFKSAMPKGYVQERTISFKKDGKY
KTRAEVKFEGRTLVNRIKLKGRDFKEKGNILGHKLRYNFNSHKVYITADKRGGSGSGSS
GGTQLLSGIVQQQNNLLRAIEAQQHLLQLTVWGIKQLQARILAGGSGGHTTWMEWDRE
INNYTSLIHSLIEESQNQQEKNEQELLEGSSGGQLLSGIVQQQNNLLRAIEAQQHLLQLTV
WGIKQLQARILAGGSGGHTTWMEWDREINNYTSLIHSLIEESQNQQEKNEQELLEGSSG
GQLLSGIVQQQNNLLRAIEAQQHLLQLTVWGIKQLQARILAGGHHHHHH

For lysate Ni-NTA pulldown assay experiments

5 Helix-His6x

(same as that used in CD)

GLUE-Cpep

MEYWHYVETTSSGQPLLREGEKDIAIDQSVGLYHGKSKILQRQRGRIFLTSQRIIYIDDAK

PTQNSLGLELDDLAYVNYSSGFLTRSPALILFFKDPSSSTEFVQLSFRKSDGVWFSWATEI

ALYTILIHSLIEESQNQQEKNEQELL

REFERENCES

1. James, A.W. and L.M. Christopher, *Reaching for high-hanging fruit in drug discovery at protein–protein interfaces.* Nature, 2007. **450**(7172): p. 1001.
2. Overington, J.P., B. Al-Lazikani, and A.L. Hopkins, *How many drug targets are there?* Nature Reviews Drug Discovery, 2006. **5**(12): p. 993-996.
3. Angelo, N.G. and P.S. Arora, *Nonpeptidic Foldamers from Amino Acids: Synthesis and Characterization of 1,3-Substituted Triazole Oligomers.* Journal of the American Chemical Society, 2005. **127**(49): p. 17134-17135.
4. Davis, J.M., L.K. Tsou, and A.D. Hamilton, *Synthetic non-peptide mimetics of α-helices.* Chemical Society Reviews, 2007. **36**(2): p. 326-334.
5. Ross, N.T., W.P. Katt, and A.D. Hamilton, *Synthetic mimetics of protein secondary structure domains.* Philosophical Transactions of the Royal Society A, 2010. **368**(1914): p. 989-1008.
6. Joshua, A.K., *Stapled peptides: Magic bullets in nature's arsenal.* Nature Chemical Biology, 2010. **6**(8): p. 566.
7. Verdine, G. and G. Hilinski, *Stapled Peptides for Intracellular Drug Targets.* Methods In Enzymology: Protein Engineering For Therapeutics, Vol 203, Pt B, 2012. **503**(1): p. 3-33.
8. Walensky, L.D. and G.H. Bird, *Hydrocarbon-stapled peptides: principles, practice, and progress.* Journal of medicinal chemistry, 2014. **57**(15): p. 6275.
9. Wang, D., et al., *Evaluation of biologically relevant short alpha-helices stabilized by a main-chain hydrogen-bond surrogate.* Journal of the American Chemical Society, 2006. **128**(28): p. 9248.

10. Catherine, M.G., et al., *Foldamers as versatile frameworks for the design and evolution of function.* Nature Chemical Biology, 2007. **3**(5): p. 252.

11. Chapman, R.N., G. Dimartino, and P.S. Arora, *A highly stable short alpha-helix constrained by a main-chain hydrogen-bond surrogate.* Journal of the American Chemical Society, 2004. **126**(39): p. 12252.

12. Gellman, S., *Foldamers: A manifesto*, in *Accounts Chem. Res.* 1998. p. 173-180.

13. Hill, D., et al., *A field guide to foldamers.* Chemical Reviews, 2001. **101**(12): p. 3893-4011.

14. Martinek, T.A. and F. Flp, *Peptidic foldamers: ramping up diversity.* Chemical Society Reviews, 2012. **41**(2): p. 687-702.

15. Azoitei, M.L., et al., *Computation-guided backbone grafting of a discontinuous motif onto a protein scaffold.* Science (New York, N.Y.), 2011. **334**(6054): p. 373.

16. Kritzer, J.A., et al., *Miniature Protein Inhibitors of the p53–hDM2 Interaction.* ChemBioChem, 2006. **7**(1): p. 29-31.

17. Liu, S., et al., *Nonnatural protein–protein interaction-pair design by key residues grafting.* Proceedings of the National Academy of Sciences of the United States of America, 2007. **104**(13): p. 5330-5335.

18. Samuel, K.S. and S.K. Peter, *Protein grafting of an HIV-1-inhibiting epitope.* Proceedings of the National Academy of Sciences of the United States of America, 2003. **100**(17): p. 9756.

19. Tlatli, R., et al., *Grafting of functional motifs onto protein scaffolds identified by PDB screening – an efficient route to design optimizable protein binders.* FEBS Journal, 2013. **280**(1): p. 139-159.

20. Chan, D.C. and P.S. Kim, *HIV Entry and Its Inhibition.* Cell, 1998. **93**(5): p. 681-684.

21. Dong, X.-N., et al., *N- and C-domains of HIV-1 gp41: mutation, structure and functions.* Immunology Letters, 2001. **75**(3): p. 215-220.

22. Buzon, V., et al., *Crystal Structure of HIV-1 gp41 Including Both Fusion Peptide and Membrane Proximal External Regions (Structure of gp41 with MPER and FPPR).* PLoS Pathogens, 2010. **6**(5): p. e1000880.

23. Chan, D.C., et al., *Core Structure of gp41 from the HIV Envelope Glycoprotein.* Cell, 1997. **89**(2): p. 263-273.

24. Cai, L., M. Gochin, and K. Liu, *Biochemistry and Biophysics of HIV-1 gp41 - Membrane Interactions and Implications for HIV-1 Envelope Protein Mediated Viral-Cell Fusion and Fusion Inhibitor Design.* Current Topics in Medicinal Chemistry, 2011. **11**(24): p. 2959-2984.

25. Eckert, D. and P. Kim, *Mechanisms of viral membrane fusion and its inhibition.* Annual Review Biochemistry, 2001. **70**: p. 777-810.

26. Lu, K., et al., *Development of HIV-1 fusion inhibitors targeting gp41.* Current medicinal chemistry, 2014. **21**(17): p. 1976.

27. Lu, L., et al., *Development of Small-molecule HIV Entry Inhibitors Specifically Targeting gp120 or gp41.* Current topics in medicinal chemistry, 2016. **16**(10): p. 1074.

28. Eckert, D.M. and P.S. Kim, *Design of Potent Inhibitors of HIV-1 Entry from the gp41 N-Peptide Region.* Proceedings of the National Academy of Sciences of the United States of America, 2001. **98**(20): p. 11187-11192.

29. Eckert, D.M., et al., *Inhibiting HIV-1 Entry: Discovery of D-Peptide Inhibitors that Target the gp41 Coiled-Coil Pocket.* Cell, 1999. **99**(1): p. 103-115.

30. Root, M. and H. Steger, *HIV-1 gp41 as a Target for Viral Entry Inhibition.* Current Pharmaceutical Design, 2004. **10**(15): p. 1805-1825.

31. Welch, B.D., et al., *Design of a Potent D-Peptide HIV-1 Entry Inhibitor with a Strong Barrier to Resistance.* The Journal of Virology, 2010. **84**(21): p. 11235.

32. Abel, S., D. Back, and M. Vourvahis, *Maraviroc: pharmacokinetics and drug interactions.* Antiviral Therapy, 2009. **14**(5): p. 607-618.

33. Lalezari, P.J., et al., *A phase II clinical study of the long-term safety and antiviral activity of enfuvirtide-based antiretroviral therapy.* AIDS, 2003. **17**(5): p. 691-698.

34. Patel, I., et al., *Pharmacokinetics, Pharmacodynamics and Drug Interaction Potential of Enfuvirtide.* Clinical Pharmacokinetics, 2005. **44**(2): p. 175-186.

35. Teo, H., et al., *ESCRT-I Core and ESCRT-II GLUE Domain Structures Reveal Role for GLUE in Linking to ESCRT-I and Membranes.* Cell, 2006. **125**(1): p. 99-111.

36. Jenssen, H. and S.I. Aspmo, *Serum stability of peptides.* Methods in Molecular Biology, 2008. **494**: p. 177-186.

37. Root, M.J., M.S. Kay, and P.S. Kim, *Protein Design of an HIV-1 Entry Inhibitor.* Science, 2001. **291**(5505): p. 884-888.

38. Blakeley, B.D., A. Chapman, and B. McNaughton, *Split-superpositive GFP reassembly is a fast, efficient, and robust method for detecting protein-protein interactions in vivo.* Molecular Biosystems, 2012. **8**(8): p. 2036-2040.

39. Chapman, Alex m. and Brian r. McNaughton, *Scratching the Surface: Resurfacing Proteins to Endow New Properties and Function.* Cell Chemical Biology, 2016. **23**(5): p. 543-553.

40. Walker, S.N., et al., *GLUE That Sticks to HIV: A Helix-Grafted GLUE Protein That Selectively Binds the HIV gp41 N-Terminal Helical Region*. ChemBioChem, 2015. **16**(2): p. 219-222.

CHAPTER THREE

Helix-Grafted Pleckstrin Homology Domains Suppress HIV-1 Infection of CD4-Positive Cells[1]

In this work, led by second year graduate student Rachel Tennyson, I assisted in gene construction and molecular cloning, as well as experimental design of helix-grafting and planning and execution of *in vitro* ELISA experiments.

3.1 Introduction

Essentially every biological process relies heavily on a cascade of protein–protein binding events.[1] The macroscopic architecture and complex molecular diversity of protein ligands or receptors often allows them to engage partners whose large surface area and spatially disperse recognition features hinder discovery of traditional small-molecule ligands. The ubiquity of such supramolecular targets underscores the need for a complementary discovery approach that produces macromolecular agents capable of specific protein recognition. Indeed, biologics constitute a rapidly expanding sector of our pharmaceutical arsenal. With an eye toward the vast array of protein–protein interfaces (PPIs) comprised of an α-helix bound into a surface cleft, we recently developed a method for helical ligand discovery conversed in Chapter 2.[2]

We began with the premise that isolated helical fragments were themselves poor starting points for either discovery or application of new ligand sequences. As discussed in detail in

[1] Reprinted with permission from:
Tennyson, R.L., Walker, S.N., Ikeda, T., Harris, R.S., Kennan, A.J., McNaughton, B.R., Helix-Grafted Pleckstrin Homology Domains Suppress HIV-1 Infection of CD4-Positive Cells, *ChemBiochem*, 2016, 17, 1892

Chapter 1, typical helix lengths at PPIs are relatively short, which complicates both production and stability of the excised sequences. Short peptides are difficult to express recombinantly, and their chemical synthesis on a therapeutically viable scale remains a significant challenge. Even when successful, it dramatically increases the cost of treatment.[3] Meanwhile, such ligands are typically unfolded, significantly enhancing their susceptibility to non-specific degradation *in vivo*.[3] These shortcomings have fueled the search for alternative structures that mimic native PPI ligands.[4-10] Native sequences have been fitted with conformational constraints (hydrogen bond surrogates, "staples") or backbone modifications (β-peptides), among other strategies. Though often successful, these approaches require nontrivial chemical synthesis that limits throughput and elevates costs. Alternatively, stably folded peptides or small proteins with intrinsic helical domains have been resurfaced (or grafted) with key contact residues from a particular PPI, although the expression and solubility of these proteins has, in some cases, been poor.[11] Though the final ligands are often still prepared synthetically, the all-natural sequences can be optimized by using directed evolution techniques.

In developing a general scaffold for helix display, we sought to identify a protein fold that presented a properly folded α-helix within the confines of a larger structure, in such a way as to permit direct receptor access to one helical face. Ideally, the basic scaffold would be readily expressible in soluble form, tolerant of mutation and/or extension of the helix (to permit sequence optimization), and protective against rapid helix degradation. Once identified, this basic framework could then serve in plug-and-play fashion as the starting point for optimizing a broad range of future PPI modulators. Our initial scaffold search identified pleckstrin homology (PH) domains as a plausible starting place. PH domains are found in a broad family of lipid-binding proteins whose fold displays a single α-helix atop a pair of β-sheets, such that one helix face is

solvent exposed.[12-14] We reasoned that such structures might serve as a good foundation for helix-grafted systems, in which solvent-exposed helix residues are mutated and/or the native helix is extended so that they mimic disease-relevant helical ligands. The reliability of expression, folded state stability, and potential for evolving optimized sequences would make such constructs excellent leads for new protein ligand therapeutics. Our initial proof-of-principle report demonstrated that a helix-grafted PH domain could replicate binding of the native helical ligand to an HIV fusion protein model.[2] Here, we explored the scope of viable PH scaffolds and generated constructs capable of inhibiting HIV infection in a live-virus assay.

3.2 Optimizing the Pleckstrin Homology Domain Scaffold for Helix-grafted Display

In the previous chapter, we described using the PH domain from a yeast-derived protein called GLUE as a scaffold for gp41 Cpeptide helix-grafted display (Figure 3.2A).[2] Surface-exposed GLUE helix residues were mutated to match those of gp41, and the helix was extended to match the native C-peptide length by using the pure gp41 sequence.[15-21] The resulting grafted protein was shown to be stable, well folded, and capable of recognizing a standard gp41 model with fidelity comparable to that of the wild-type peptide, as measured by split-GFP reassembly, ELISA, and copurification assays. It also retained much of its efficacy even after a 12 h exposure to human serum, supporting our initial design hypothesis that positioning the ligand within a larger, stably folded structure would protect against rapid degradation.[19, 20] Although our initial effort was quite successful, we were interested both in probing the scope of our chosen scaffold and in testing our gp41 mimics in more demanding environments. We began by selecting a collection of nine PH domains, favoring those of human origin to potentially mitigate downstream immunogenicity. The set of PH domains with reported X-ray structures exhibited considerably

diversity in helix length, total protein size, and percentage of total residues involved in the helix. As we were unsure how each of these variables might impact expression, viability and/or helix presentation, we chose an array of candidates covering a range of characteristics (Figure 3.2B, Table 3.2). Although the canonical PH fold presents a C-terminal helix, the engulfment and cell motility (ELMO)-1 domain has the added benefit of an additional N-terminal helix,[22] providing the option to display partially structured ligands that are helical at both termini.

Our initial screen for scaffold suitability was simple expression of the wild-type proteins, as well-expressing systems were expected to better tolerate modification. At this stage, four of our nine next generation candidates were eliminated based on poor expression (Figure 3.2C and Supplemental Figure 6.2A). The remaining candidates, including our original GLUE scaffold, were grafted with the gp41 Cpeptide sequence. We tested both the N- and C-terminal ELMO sites (Figure 3.2D) and were satisfied to discover that both constructs expressed admirably (Figure 3.2E and Supplemental Figure 6.2A). Although other sequences also worked, ELMO seem to present the most flexibility and thus the most potential to serve as a general scaffold for a series of future helical ligands. We therefore focused on ELMO as our platform of choice moving forward.

Table 3.2 Candidate PH Domains with a range of characteristics from protein size to percent helicity. Many of the selected scaffolds were of human origin to mitigate potential immunogenicity effects as a therapeutic.

Table 3.2: Candidate PH Domains

Domains	PDB	# of residues			human	disulfide
		helix	total	% helicity		
AKT2	1P6S	20	111	18	Yes	Yes
APPL1 PH	2ELB	18	112	16	Yes	No
APPL1 PTB	2ELA	36	145	25	Yes	No
DOK5	1J0W	21	104	20	Yes	No
DYNAMIN	2DYN	13	105	12	Yes	No
ELMO (C-term)	2VSZ	18	147	12	Yes	No
ELMO (N- term)	2VSZ	27	147	19	Yes	No
GLUE	2CAY	18	128	14	Yeast	No
HOMER	1I2H	26	116	22	Rat	No
PLECKSTRIN	2I5F	17	105	16	Yes	No
PKB	1UNP	30	119	25	Yes	Yes

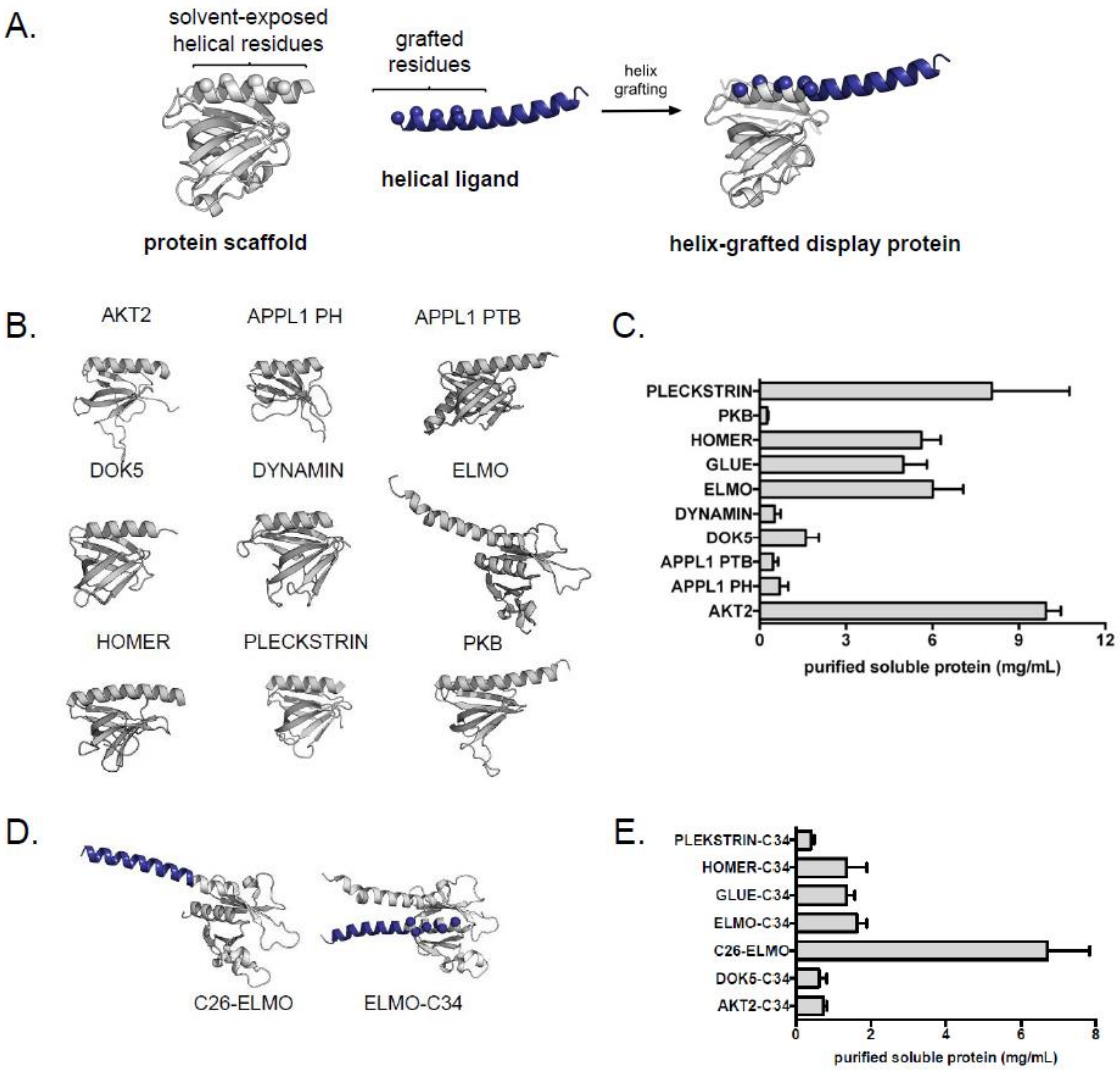


Figure 3.2 (A) Helix-grafted display strategy for generating stable Cpeptide mimics on a PH domain. The native ligand (density blue, PDB: 1AIK) is overlaid on the scaffold protein (grey, PDB: 2CAY), and solvent-exposed scaffold residues are mutated to those of the displayed ligand (density blue spheres). (B) Candidate PH domain scaffolds examined in this work. PDB: 1P6S (AKT2), 2ELA (APPL1 PTB), 2ELB (APPL1 PH), 1J0W (DOK5), 2DYN (dynamin), 1I2H (homer), 2VSZ (ELMO), 1UNP (PKB), 2I5F (Plekstrin). (C) Expression levels of soluble wildtype scaffolds following purification by His6/Ni-NTA column. Error bars indicate standard deviation of three experiments. (D) Models of N-terminal (C26-ELMO) and C-terminal (ELMO-C34) helix grafts on the ELMO scaffold. Grafted gp41 positions are density blue (spheres indicate mutated α-carbon positions). Soluble expression levels (following purification by His6/Nickel-NTA column purification).

3.3 Optimizing the Cpeptide Helical Graft

Having centered on ELMO as our most promising display vehicle, we sought to test its ability to effectively present a collection of different Cpeptide sequences, while at the same time probing the sensitivity of our grafted constructs to alterations in ligand sequence. Although the C34 peptide used in our original helix-grafted ligand is taken from a partially overlapping region of gp41 as Enfuvirtide (T20), a significant portion of each sequence is unique. We thus sought to test grafting capabilities of both sequences, in addition to the longer one formed by their union (C46, Table 3.3). To test the intrinsic flexibility of N- versus C-terminal display, we prepared ELMO scaffolds grafted at either terminus (Lig-ELMO or ELMO-Lig), with C34, T20, and C46. We were especially interested in probing the role of the N-terminal WWI triad on C34, the side chains of which bind into a deep and highly conserved hydrophobic groove on the NHR trimer surface. The absence of this interaction in T20 (the sequence of which is shifted more toward the C-peptide C terminus) perhaps reduces barriers to acquired resistance (Table 3.3).

Table 3.3 Sequence of grafted gp41 Cpeptide ligand region for N- and C-terminal helix-grafted ELMO (designated lig-ELMO and ELMO-lig, respectively).

Table 3.3: Candidate Cpeptide Grafts

helix residue	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
C26-ELMO	W	M	E	W	D	R	E	I	N	N	Y	T	S	L	I	H	S	L	I	E	E	S	Q	N	Q
C34-ELMO	W	M	E	W	D	R	E	I	N	N	Y	T	S	L	I	H	S	L	I	E	E	S	Q	N	Q
T20-ELMO											Y	T	S	L	I	H	S	L	I	E	E	S	Q	N	Q
C46-ELMO	W	M	E	W	D	R	E	I	N	N	Y	T	S	L	I	H	S	L	I	E	E	S	Q	N	Q
ELMO-C34	W	E	Y	W	I	W	T	I	G	L	Y	T	L	L	G	K	S	L	I	E	E	S	Q	N	Q
ELMO-T20											Y	E	Y	L	I	W	T	L	G	L	E	S	L	L	G
ELMO-C46	W	E	Y	W	I	W	T	I	G	L	Y	T	L	L	G	K	S	L	I	E	E	S	Q	N	Q

helix residue	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46
C26-ELMO	Q																				
C34-ELMO	Q	E	K	N	E	Q	E	L	L												
T20-ELMO	Q	E	K	N	E	Q	E	L	L	E	L	D	K	W	A	S	L	W	N	W	F
C46-ELMO	Q	E	K	N	E	Q	E	L	L	E	L	D	K	W	A	S	L	W	N	W	F
ELMO-C34	Q	E	K	N	E	Q	E	L	L												
ELMO-T20	K	E	K	N	E	Q	E	L	L	E	L	D	K	W	A	S	L	W	N	W	F
ELMO-C46	Q	E	K	N	E	Q	E	L	L	E	L	D	K	W	A	S	L	W	N	W	F

The C-terminal grafts were prepared as before, with five solvent-exposed positions on the native ELMO helix mutated corresponding to those of the HIV sequence, and the C-terminus of the protein extended by the appropriate amount by using the remaining gp41 sequence. In contrast

to the C-terminal ELMO helix, which rests atop the β-sheets, the N-terminal helix is more fully exposed and of considerable length (nearly 20 residues). Fearing that long extensions of this helix might not be well tolerated, we prepared C34-ELMO and T20-ELMO by first excising 17 residues from the native helix and then simply fusing the appropriate sequence to the N terminus of the resulting protein (termed Δ17-ELMO). For C46-ELMO, three additional residues were trimmed, as the fused helix was longer (termed Δ21-ELMO). Finally, as a hedge against even those shorter helices proving unworkable, we prepared C26-ELMO, in which eight residues were deleted from the C-terminus of C34 before fusing it to the shortened Δ17-ELMO scaffold. To determine the impact of the N-terminal deletions from the core ELMO protein, we prepared both Δ17-ELMO and Δ21-ELMO on their own and were pleased to discover that both expressed even better than the wild-type platform (Figure 3.3A), so much so that we used Δ21-ELMO as the starting point for the C-terminal ELMO-Lig constructs.

To establish that these various modifications had not compromised the overall protein fold, we examined each grafted protein by circular dichroism (CD) spectroscopy and compared the results to those of the base proteins (wild-type, Δ17-, and Δ21-ELMO). Plots of mean residue ellipticity versus wavelength for the C-terminal fusions were consistent with retention of overall structure and displayed expected helicity increases compared to the appropriate control. Similarly, all of the N-terminal fusion proteins exhibited increased helicity with respect to each of the starting proteins (Figure 3.3B). The helix-grafted proteins also displayed similar thermal unfolding profiles, consistent with retention of macroscopic structure (Figure 3.3C). Having determined that the HIV ligand mimic candidates were behaving as expected, we moved on to evaluate their viability as ligands for both model and live virus receptors.

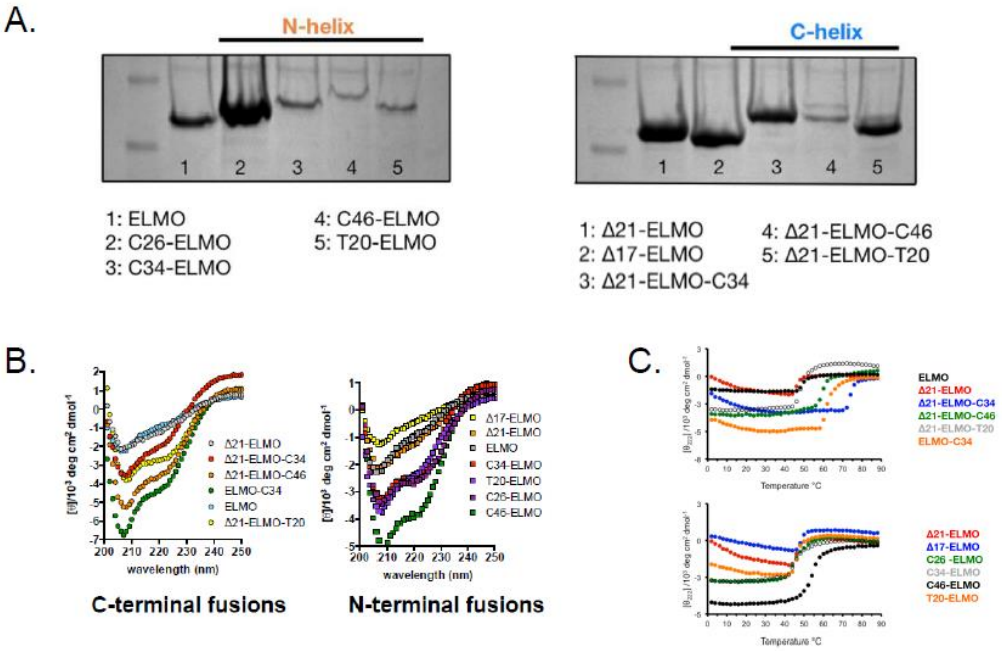


Figure 3.3 (A) SDS-PAGE gels after Ni-NTA column purification indicating protein expression levels of N- and C-terminal grafts to ELMO. The wildtype ELMO scaffold, Δ17-ELMO, and Δ21-ELMO scaffolds were ran to compare protein expression of helical grafts to the protein scaffolds. (B) Circular dichroism spectra of C- and N-terminal ELMO grafts. (C) Circular dichroism thermal melt data at 222 nm for N- and C-terminal grafts to ELMO, showing temperature-dependent unfolding with similar theral stability as compared to the original protein scaffolds.

3.4 Analysis of Optimized Helix-grafted Display Proteins

As in the previous chapter, initial evaluations of ligand–receptor binding employed the 5-helix construct that has been widely used for this purpose.[23, 24] The 5-helix protein links the central coiled-coil NHR trimer of gp41 and two CHR ligands into one sequence that assembles to present a single C-peptide binding site. This strategy both simplifies the interaction to be measured (making it a 1:1 rather than 3:1 complex) and eases solubility issues arising from exposing three copies of the significantly hydrophobic ligand-binding surface.

We initially performed an *in cellulo* ELISA assay to assess binding in a complex cellular environment. Somewhat surprisingly, the native N-helix on ELMO, which we originally feared

might compromise expression efficiency, also had some nonspecific affinity for the 5-helix protein in this assay. However, this non-specific binding could be abrogated by truncation of the N-terminal helix. Two shortened variants, Δ17-ELMO and Δ21-ELMO, with 17 or 21 residues removed from the N-terminal helix, respectively, showed substantially decreased affinity for the 5-helix protein in the ELISA assay. The N-terminal fusion proteins generated stronger signals, particularly those with the WWI residues; C26-ELMO and C34-ELMO were better than C46-ELMO, which in turn was better than T20-ELMO (Figure 3.4A). The two shorter fusion proteins gave signals comparable to the native C34 peptide. The C-terminal fusion proteins were weaker, with only Δ21-ELMO-C34 giving a signal even comparable to the analogous control scaffold (Δ21-ELMO).

We corroborated these *in cellulo* findings using a nickel pulldown assay. Cells were transformed and expressed with a plasmid encoding both the untagged ligand candidate and His-tagged 5-helix protein, lysed, and exposed to Ni-NTA agarose resin. Consistent with the data from our ELISA experiment, SDS-PAGE analysis of the N-terminal fusions (Figure 3.4B, lanes 1–6) suggested that those bearing the WWI triad (lanes 3–5) were most effective at binding to the 5-helix protein. The two shorter helices (C26 and C34) displayed the strongest retention, with a significant reduction for the longer C46-ELMO construct, and very little observable retention for T20-ELMO. The C-terminal fusion proteins (Figure 3.4B, lanes 7–12) exhibited a similar trend, albeit with an overall reduction in apparent affinity. Of the proteins with the native N-helix removed (denoted Δ21-ELMO-Lig), the C34 fusion (lane 10) was again more effective than either the C46 or T20 one (lanes 11–12), though the intact ELMO-C34 protein (lane 9) is perhaps even better due to the presence of the wildtype N-terminal helix which exhibits modest affinity for the

5-helix. Neither ungrafted truncation (Δ17- or Δ21-ELMO) bound 5-helix appreciably in the copurification assay (Figure 3.4B).

To better understand the complex involving 5-helix and our best-performing binders (C26-ELMO and C34-ELMO), we used *in vitro* ELISA with biotin-5helix-his immobilized (Supplemental Figure 6.2B) to measure their dissociation constants (K_D). Satisfyingly, these proteins tightly bound 5-helix (K_D ~91 +/-8 nM and 6 +/-2 nM, respectively, Figure 3.4C). In addition to tightly binding the 5-helix protein, C26-ELMO and C34-ELMO were relatively stable in proteolytically active human serum. For both proteins, >50% of the full-length species was detected by Western blot after incubating with proteolytically active human serum for 12 h (Figure 3.4D).

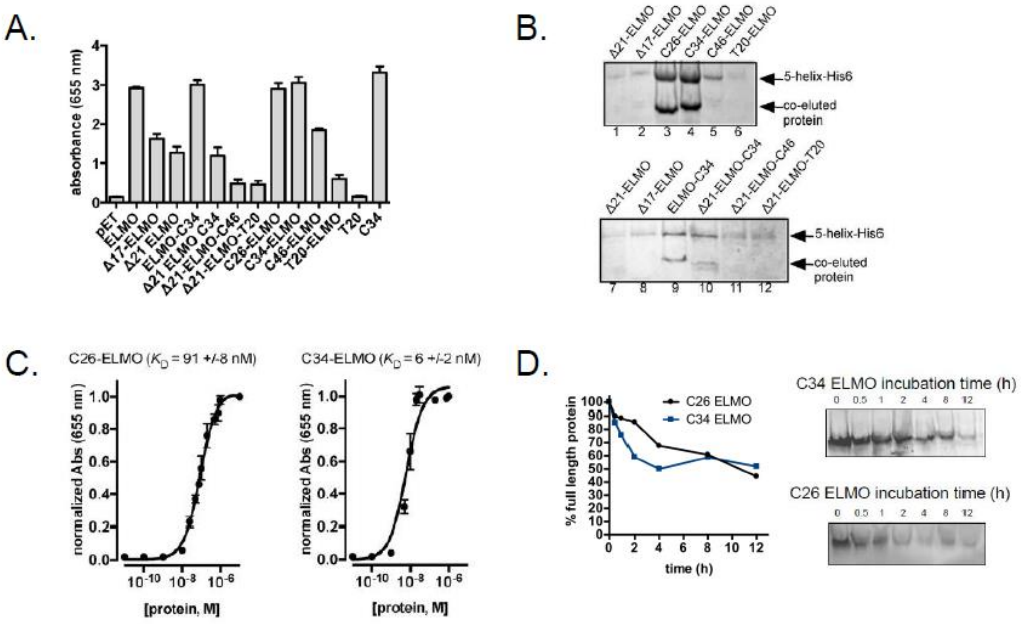


Figure 3.4 (A) *in cellulo* ELISA data showing binding between 5-helix-His_6 and ELMO derived grafted proteins. (B) Ni-NTA co-purification of 5-Helix-His_6 and either N-terminal (top) or C-terminal (bottom) helix-grafted ELMO-derived proteins. (C) Dissociation constant (K_D) for complexes involving the 5-helix protein and C26-ELMO or C34-ELMO (D) Western blot and densitometry analysis of C26-ELMO or C34-ELMO prior to human serum incubation (lane 1, time 0) and throughout incubation with human serum for up to 12 h.

3.5 HIV-1 Inhibition Using Optimized Helix-grafted Display Proteins

Collectively, these initial experiments supported the notion that C-peptide grafted ELMO scaffolds were capable of recognizing the 5-helix model system, even in complicated cellular contexts, and further suggested that the most efficient of these systems were the shorter N-terminal fusions bearing the key WWI residues. Encouraged by these data, we set out to test the ability of helix-grafted gp41 mimics to inhibit infection by actual virus, according to a previously reported protocol.[25, 26] HIV-1 IIIB was administered to CD4-positive mammalian cells stably integrated with a plasmid that encodes the HIV-1 long-terminal repeat (LTR) upstream of green fluorescent protein (GFP). Thus, if HIV-1 successfully infects these cells, HIV-1 Tat/TAR-dependent transcription ultimately leads to GFP expression. The percentage of cells that express GFP (measured by flow cytometry) is thus equivalent to the percentage infected by the virus. Consistent with the ELISA and co-purification experiments, the N-terminal fusion proteins proved more effective by this measure as well (Figure 3.5A). Only the three WWI-containing helices (C26-, C34-, and C46-ELMO) exhibited significant inhibition, with the T20-ELMO protein proving largely ineffective. The longer C46 construct again lagged behind those with shorter helices, despite containing the WWI triad. The C-terminal fusion proteins (Figure 3.5B), as before, were less efficient overall, and only Δ21-ELMO-C34 was comparable to any of the N-terminal species. The scaffold controls (Δ17- and Δ21-ELMO) did not materially inhibit infection, despite ELISA data suggesting at least moderate 5-helix affinity. Although both N- and C-terminal T20 fusions were largely ineffective in each of these assays, the isolated T20 peptide itself was significantly more potent in the live virus assay than any of our grafted systems. However, this *in vitro* assay is blind to the rapid degradation which compromises the *in vivo* efficacy of the T20 peptide and its variants, such as C34.

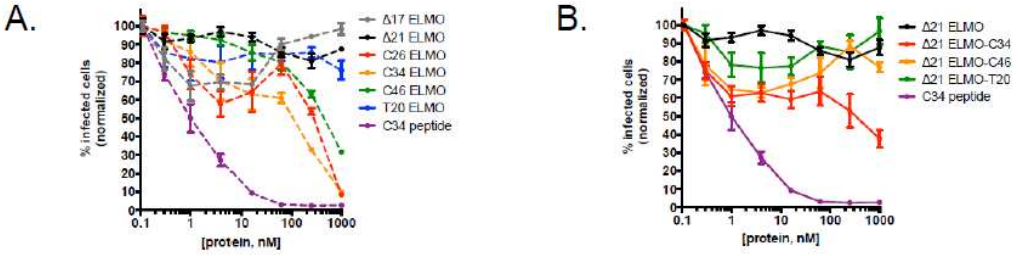


Figure 3.5 (A) Suppression of HIV-1 entry by N-terminal fusion proteins and controls Δ17-ELMO (gray); Δ21-ELMO (black); C26-ELMO (red); C34-ELMO (orange); C46-ELMO (green); T20-ELMO (blue); and C34 peptide (purple). (B) Suppression of HIV-1 entry by ELMO-derived C-terminal fusion proteins and controls: Δ21-ELMO (black); Δ21-ELMO-C34 (red); Δ21-ELMO-C46 (orange); and Δ21-ELMO-T20 (green).

3.6 Conclusion

Taken together, these results demonstrate the viability of our helix-grafting strategy as a method for presenting gp41 Cpeptide helices in a manner that allows them to efficiently recognize their intended receptor. Despite this success, opportunities for improvement remain. One plausible explanation for reduced efficacy of the longer C46 constructs (compared to C26/C34) is steric crowding, especially in the more complex *in cellulo* experiments with proximal lipid bilayers. It could be that further reduction in scaffold size is required for true generality moving forward. Additional affinity might also be obtained through directed evolution, optimizing the choice of receptor-facing residues. In the longer term, we are confident that the general strategy of helix-grafted ligand display platforms will prove amenable to the discovery and optimization of new PPI ligands for many different targets.[27]

3.7 Methods

Protein Expression and Purification

Genes were cloned into pET using restriction enzymes BamHI and KpnI, downstream of a His6x tag and transformed into BL21s (DE3). Proteins containing a disulfide bond were transformed into Shuffle T7s. Cells were grown in 0.5 L LB cultures containing 50 mg/mL carbenicillin at 37 °C to OD_{600} =0.5 - 0.6 and induced with 1 μM IPTG at 25 °C overnight. Cells were then collected by centrifugation, resuspended in Tris buffer (20 mM Tris pH 7.4, 100 mM NaCl, 10 mM $(NH_4)_2SO_4$) and stored at -20 °C. Frozen pellets were thawed and sonicated with 1 second pulses for 2 minutes. The lysate was cleared by centrifugation (8,000 rpm 10 min.) and the supernatant was mixed with 1 mL of Ni-NTA agarose resin for 30 min at 4 °C. The resin was collected by centrifugation (4950 rpm, 10 min.). The resin was sequentially washed with 50 mL of Tris buffer containing 20 mM imidazole, 50 mL buffer containing 50 mM imidazole, and 10 mL Tris buffer containing 75 mM imidazole. The protein was then eluted with 2 mL Tris buffer containing 400 mM imidazole. The proteins were dialyzed against Tris buffer and analyzed for purity by SDS-PAGE. Purified protein concentrations were quantified using Beer's Law at an absorbance of 280 nm, following standard practice.

Resolubilization of 5-Helix Inclusion Bodies

5Helix-His6x was cloned into a modified pETDuet-1 vector using restriction enzymes NdeI and KpnI and transformed into BL21s (DE3). Cells were induced to express 5 Helix-His6x and lysed as described above. The lysate was cleared by centrifugation (9,500 rpm, 30 min.) and the supernatant discarded. The pellet was washed twice with Tris buffer containing 0.5 % Triton® X-100 and once with Tris buffer. The pellet was resuspended in urea buffer (Tris buffer with 8 M urea and 10 mM imidazole) to resolubilize the inclusion bodies and cleared by centrifugation

(9,500 rpm, 30 min.) The supernatant was mixed with 1 mL of Ni-NTA agarose resin for 1 hour at 4 °C. The resin was collected by centrifugation (4950 rpm, 4 min.). The resin was washed with 50 mL of urea buffer and eluted with 40 mL of urea elution buffer (Tris buffer with 6 M urea and 100 mM imidazole) into 500 mL Tris buffer by gravity dripping while stirring to refold the protein. The 540 mL elution was run through a column containing 1 mL of Ni-NTA agarose resin and eluted with 5 mL Tris buffer containing 400 mM imidazole. The protein was then eluted with 4 mL buffer containing 400 mM imidazole. The proteins were dialyzed against buffer and analyzed for purity by SDS-PAGE and analyzed for refolding by CD.

Circular Dichroism

Proteins were purified as described above. Separately, each protein was diluted to 5-12 μM in Tris buffer. Wavelength data are the average of three scans from 250 nm to 200 nm in 1 nm steps at 25 °C. Thermal denaturation experiments at 222 nm were run from 0 to 90 °C in two-degree steps at a two-degree/minute rate of increase with one-minute equilibration and data averaging at each temperature. Tm values were obtained from minima of the first derivative of θ versus 1/T plots.

Serum Stability

The helix-grafted C26-ELMO protein and helix-grafted C34-ELMO protein were cloned into a pET vector upstream of a C-terminal His6x-GS-FLAG tag (DYKDDDDK) using restriction enzymes NheI and BamHI. Proteins were purified as described above in Phosphate buffer (20 mM Na_2HPO_4 pH 7.4, 100 mM NaCl). 1 mL of RPMI supplemented with 25 % (v/v) of human serum was equilibrated at 37°C. C26-ELMO and C34-ELMO were each added to the solution to obtain a final concentration of 50 μg/mL and incubated at 37 °C. At known time intervals of 0.5, 1, 2,4, 8, 12, 24 hours, 100 μL of the reaction solution was removed and denatured at 94 °C for 20 minutes and stored at -80 °C. Samples were separated by SDS-PAGE and transferred to a nitrocellulose

membrane via an iBlot western blotting apparatus. The membrane was washed with TRIS and then incubated in Blocking Buffer at room temperature for 1 hour. The membrane was subsequently washed 3x with TRIS containing 0.1 % Tween® 20. The membrane was then incubated with a mouse anti-DDDDK antibody (antiFLAG) in primary antibody wash buffer (TRIS with 5% BSA and 0.1 % Tween® 20) for 3 hour at 4 °C. The membrane was then washed 3x with TRIS containing 0.1 % Tween® 20, and then incubated with a IRDye 800CW Goat anti-mouse IgG secondary antibody in Blocking Buffer for 1 hour at room temperature. The membrane was washed with PBS containing 0.1 % Tween-20 and imaged using the Odyssey Classic Infrared Imager.

Cell lysate ELISA

Each helix-grafted ELMO protein, T20, and C34 were cloned into MCS1 of pETDuet-1 with FLAG tags using restriction enzymes NcoI and NotI. The 5-Helix with a C-terminal His6x tag was cloned into MCS2 of pETDuet-1 using restriction enzymes NdeI and KpnI. Completed constructs were transformed into BL21s (DE3). Cells containing the co-expressed pair were inoculated and induced in 10 mL LB cultures overnight. Cells were spun down and resuspended in 10 mL buffer (20 mM Tris pH 7.4, 100 mM NaCl, 10 mM $(NH_4)_2SO_4$), lysed by sonication, and spun down to remove cell debris. Cleared lysates were incubated on clear Ni-NTA coated plates for 1 hr. at room temperature and washed 3x with 200 µL wash buffer (20 mM Tris pH 7.4, 100 mM NaCl, 10 mM $(NH_4)_2SO_4$, 0.1% Tween® 20, 0.01 mg/mL BSA). HRP-conjugated mouse anti-DDDDK antibody in LICOR Blocking Buffer was incubated for 1 hr at room temperature, followed by 3x 200 µL washes (5 min.). Color was developed using TMB-One substrate and absorbance was measured at 655nm on a SynergyMx Microplate Reader.

Lysate Ni-NTA Pulldown Assay

Each helix grafted ELMO protein was cloned into MCS1 of pETDuet-1 using restriction enzymes NcoI and NotI. The 5-helix with a C-terminal His6x tag was cloned into MCS2 of pETDuet-1 using the restriction enzymes NdeI and KpnI. Completed constructs were transformed into BL21s (DE3). Cells containing the co-expressed pair were inoculated and induced in 100 mL LB cultures overnight. Cells were spun down and resuspended in 5 mL buffer (100 mM NaCl, 20 mM Tris pH 7.4, 10 mM $(NH_4)_2SO_4$, lysed by sonication, and spun down to remove cell debris. Cleared lysate was incubated with 300 μL Ni-NTA agarose resin for 30 min. The resin was collected by centrifugation (4950 rpm, 10 min.). The resin was sequentially washed with 10 mL of Tris buffer containing 50 mM imidazole and 5 mL Tris buffer containing 100 mM imidazole. Proteins were eluted with 500 μL Tris buffer containing 400 mM imidazole. The pulldown was analyzed by SDS-PAGE.

Biotinylation

5helix-His6x was cloned into a pET vector containing an upstream Avitag™ (GLNDIFEAQKIEWHE)-GGSGGSGGT linker using restriction enzymes KpnI and PacI. The protein was resolublized from inclusion bodies as described above in Phosphate buffer (20 mM Na_2HPO_4 pH 7.4, 100 mM NaCl). His6x-BirA was cloned into pET using restriction enzymes NcoI and KpnI and purified as described above in Phosphate buffer. 300 mL of Avitagged™ 5-Helix-His6x protein at 38 μM was incubated with 6 mL of His6x-BirA at 1 mg/mL using Avidity® BirA biotin-protein ligase standard reaction kit at 30°C for 40 min. Biotinylation was confirmed by Agilent 6220 TOF LC-MS.

In vitro ELISA

The helix-grafted C26-ELMO protein and helix-grafted C34-ELMO protein were cloned into a pET vector upstream of a C-terminal His6-GS-FLAG tag (DYKDDDDK) using restriction enzymes NheI and BamHI. Proteins were purified as described above in Phosphate buffer (20 mM Na_2HPO_4 pH 7.4, 100 mM NaCl). The biotinylated Avitagged 5-helix-His6 (biotin-5helix) was prepared as described above in Phosphate buffer and diluted to 10 μg/mL. Pierce® Streptavidin-coated clear 96-well plates with a binding capacity of 5 pmol were pre-blocked with 200 μL of wash buffer (Phosphate buffer, 0.5 mg/mL BSA, 0.1% tween® 20) for 1 hour. Biotin-5helix was immobilized on the streptavidin-coated plates by incubating 100 μL of diluted protein for 1 hour at room temperature, followed by 4x 200 μL washes (5 min). 100 mL of C26-ELMO or C34-ELMO was incubated in various concentrations (10 pM, 100 pM, 1 nM, 5 nM, 10 nM, 20 nM, 25 nM, 30 nM, 50 nM, 75 nM, 100 nM, 200 nM, 500 nM, 750 nM, 1 μM, 10 μM) for 1 hour, followed by 4x 200 μL 4°C washes (5 min). (All concentrations of incubated helix-grafted protein were above ligand-depleting conditions for binding the 5 pmol immobilized biotin-5helix in each well). A 1:10,000 dilution of HRP-conjugated mouse anti-DDDDK antibody in LICOR Blocking Buffer was incubated for 1 hour at room temperature, followed by 4x 200 μL 4°C washes (5 min). Color was developed using TMB-One substrate and absorbance was measured at 655nm on a SynergyMx Microplate Reader.

Infectivity Assay

HEK293T cells were maintained in DMEM supplemented with 10% fetal bovine serum (FBS) and 0.5% penicillin/streptomycin (P/S). CEM-GFP cells were cultured in RPMI with 10% FBS and 0.5% P/S. The procedure was virtually identical to a previously published variant. The HIV-1 IIIB C200 proviral expression construct has been described previously. Viruses were produced by

transfection of 3.0 μg of Vif-proficient proviral expression construct into 293T cells ($3.0x10^6$) using TransIT®-LT1 reagent. 48hr later, virus-containing supernatants were filtered by 0.45 μm filters and used to infect into $2.5x10^4$ CEM-GFP cells with varying concentration of inhibitors. Infectivity (GFP+ cells) was measured by flow cytometry at 2 days post-infection.

3.8 Sequences of Proteins Used in this Work

His6x-PH domain scaffold-native helix-grafted C34 residues

His6-wtGLUE

MGSSHHHHHHSQDPEYWHYVETTSSGQPLLREGEKDIFIDQSVGLYHGKSKILQRQRGR
IFLTSQRIIYIDDAKPTQNSLGLELDDLAYVNYSSGFLTRSPALILFFKDPSSSTEFVQLSFR
KSDGVLFSQATERALENILT

His6-GLUE-C34

MGSSHHHHHHSQDPEYWHYVETTSSGQPLLREGEKDIFIDQSVGLYHGKSKILQRQRGR
IFLTSQRIIYIDDAKPTQNSLGLELDDLAYVNYSSGFLTRSPALILFFKDPSSSTEFVQLSFR
KSDGVWFSWATEIALYTILIHSLIEESQNQQEKNEQELL

His6-wtAKT2

MGSSHHHHHHSQDPMNEVSVIKEGWLHKRGEYIKTWRPRYFLLKSDGSFIGYKERPEA
PDQTLPPLNNFSVAECQLMKTERPRPNTFVIRCLQWTTVIERTFHVDSPDEREEWMRAIQ
MVANSLK

His6-AKT2-C34

MGSSHHHHHHSQDPMNEVSVIKEGWLHKRGEYIKTWRPRYFLLKSDGSFIGYKERPEA
PDQTLPPLNNFSVAECQLMKTERPRPNTFVIRCLQWTTVIERTFHVDSPWERWEWMIAI
YTVAIHSLIEESQNQQEKNEQELL

His6-wtAPPL1 PH

MGSSHHHHHHSQDPVPDPDPTKFPVNRNLTRKAGYLNARNKTGLVSSTWDRQFYFTQ

GGNLMSQARGDVAGGLAMDIDNCSVMAVDCEDRRYCFQITSFDGKKSSILQAESKKDH

EEWICTINNISKQ

His6-wtAPPL1 PTB

MGSSHHHHHHSQDPFIVRFLGSMEVKSDDHPDVVYETMRQILAARAIHNIFRMTESHLL

VTCDCLKLIDPQTQVTRLTFPLPCVVLYATHQENKRLFGFVLRTSSGRSESNLSSVCYIFE

SNNEGEKICDSVGLAKQIALHAELDRRASEKQKEIERVK

His6-wtDOK5

MGSSHHHHHHSQDPREQSERFNVYLMPSPNLDVHGECALQITYEYICLWDVQNPRVKLI

SWPLSALRRYGRDTTWFTFEAGRMCETGEGLFIFQTRDGEAIYQKVHSAALAIAELER

His6-DOK5-C34

MGSSHHHHHHSQDPREQSERFNVYLMPSPNLDVHGECALQITYEYICLWDVQNPRVKLI

SWPLSALRRYGRDTTWFTFEAGRMCETGEGLFIFQTRDGEAIYQKVHWAAWDIAINNY

TSLIHSLIEESQNQQEKNEQELL

His6-wtDYNAMIN

MGSSHHHHHHSQDPILVIRKGWLTINNIGIMKGGSKEYWFVLTAENLSWYKDDEEKEK

KYMLSVDNLKLRDVEKGFMSSKHIFALFNTEQRNVYKDYRQLELACETQEEVDSWKAS

FLR

His6-wtHOMER

MGSSHHHHHHSQDPGEQPIFSTRAHVFQIDPNTKKNWVPTSKHAVTVSYFYDSTRNVY

RIISLDGSKAIINSTITPNMTFTKTSQKFGQWADSRANTVYGLGFSSEHHLSKFAEKFQEF

KEAARLAKEKSQEK

His6-HOMER-C34

MGSSHHHHHHSQDPGEQPIFSTRAHVFQIDPNTKKNWVPTSKHAVTVSYFYDSTRNVY

RIISLDGSKAIINSTITPNMTFTKTSQKFGQWADSRANTVYGLGFSSEHHLSKFAWKFWE

FKIAAYLAKIKSLEKESQNQQEKNEQELL

His6-wtPKB

MGSSHHHHHHSQDPDVAIVKEGWLHKRGEYIKTWRPRYFLLKNDGTFIGYKERPQDVD

QREAPLNNFSVAQCQLMKTERPRPNTFIIRCLQWTTVIERTFHVETPEEREEWTTAIQTV

ADGLKKQEEEEMDFR

His6-wtPLECKSTRIN

MGSSHHHHHHSQPGVIIKQGCLLKQGHRRKNWKVRKFILREDPAYLHYYDPAGAEDPL

GAIHLRGCVVTSVESNSNGRKSEEENLFEIITADEVHYFLQAATPKERTEWIKAIQMASR

TGKD

His6-PLECKSTRIN-C34

MGSSHHHHHHSQDPGVIIKQGCLLKQGHRRKNWKVRKFILREDPAYLHYYDPAGAEDP

LGAIHLRGCVVTSVESNSNGRKSEEENLFEIITADEVHYFLQAATPWERWEWIIAIYMASI

TGLIEESQNQQEKNEQELL

His6-wtELMO

MGSSHHHHHHSQDPPILELKEKIQPEILELIKQQRLNRLVEGTCFRKLNARRRQDKFWYC

RLSPNHKVLHYGDLEESPQGEVPHDSLQDKLPVADIKAVVTGKDCPHMKEKGALKQN

KEVLELAFSILYDSNCQLNFIAPDKHEYCIWTDGLNALLGK

His6-C26-ELMO

MGSSHHHHHHSQDPWMEWDREINNYTSLIHSLIEESQNQQKQQRLNRLVEGTCFRKLN

ARRRQDKFWYCRLSPNHKVLHYGDLEESPQGEVPHDSLQDKLPVADIKAVVTGKDCPH

MKEKGALKQNKEVLELAFSILYDSNCQLNFIAPDKHEYCIWTDGLNALLGK

His6-C34-ELMO

MGSSHHHHHHSQDPWMEWDREINNYTSLIHSLIEESQNQQEKNEQELLKQQRLNRLVE

GTCFRKLNARRRQDKFWYCRLSPNHKVLHYGDLEESPQGEVPHDSLQDKLPVADIKAV

VTGKDCPHMKEKGALKQNKEVLELAFSILYDSNCQLNFIAPDKHEYCIWTDGLNALLG

K

His6-C46-ELMO

MGSSHHHHHHSQDPWMEWDREINNYTSLIHSLIEESQNQQEKNEQELLELDKWASLWN

WFRLNRLVEGTCFRKLNARRRQDKFWYCRLSPNHKVLHYGDLEESPQGEVPHDSLQD

KLPVADIKAVVTGKDCPHMKEKGALKQNKEVLELAFSILYDSNCQLNFIAPDKHEYCIW

TDGLNALLGK

His6-T20-ELMO

MGSSHHHHHHSQDPYTSLIHSLIEESQNQQEKNEQELLELDKWASLWNWFKQQRLNRL

VEGTCFRKLNARRRQDKFWYCRLSPNHKVLHYGDLEESPQGEVPHDSLQDKLPVADIK

AVVTGKDCPHMKEKGALKQNKEVLELAFSILYDSNCQLNFIAPDKHEYCIWTDGLNAL

LGK

His6-ELMO-C34

MGSSHHHHHHSQDPPILELKEKIQPEILELIKQQRLNRLVEGTCFRKLNARRRQDKFWYC

RLSPNHKVLHYGDLEESPQGEVPHDSLQDKLPVADIKAVVTGKDCPHMKEKGALKQN

KEVLELAFSILYDSNCQLNFIAPDKWEYWIWTIGLYTLLGKSLIEESQNQQEKNEQELL

His6-Δ21-ELMO

MGSSHHHHHHSQDPLNRLVEGTCFRKLNARRRQDKFWYCRLSPNHKVLHYGDLEESP

QGEVPHDSLQDKLPVADIKAVVTGKDCPHMKEKGALKQNKEVLELAFSILYDSNCQLN

FIAPDKHEYCIWTDGLNALLGK

His6 -Δ17-ELMO

MGSSHHHHHHSQDPKQQRLNRLVEGTCFRKLNARRRQDKFWYCRLSPNHKVLHYGDL

EESPQGEVPHDSLQDKLPVADIKAVVTGKDCPHMKEKGALKQNKEVLELAFSILYDSNC

QLNFIAPDKHEYCIWTDGLNALLGK

His6-Δ21-ELMO-C34

MGSSHHHHHHSQDPLNRLVEGTCFRKLNARRRQDKFWYCRLSPNHKVLHYGDLEESP

QGEVPHDSLQDKLPVADIKAVVTGKDCPHMKEKGALKQNKEVLELAFSILYDSNCQLN

FIAPDKWEYWIWTIGLYTLLGKSLIEESQNQQEKNEQELL

His6-Δ21-ELMO-C46

MGSSHHHHHHSQDPLNRLVEGTCFRKLNARRRQDKFWYCRLSPNHKVLHYGDLEESP

QGEVPHDSLQDKLPVADIKAVVTGKDCPHMKEKGALKQNKEVLELAFSILYDSNCQLN

FIAPDKWEYWIWTIGLYTLLGKSLIEESQNQQEKNEQELLELDKWASLWNWF

His6- Δ21-ELMO-T20

MGSSHHHHHHSQDPLNRLVEGTCFRKLNARRRQDKFWYCRLSPNHKVLHYGDLEESP

QGEVPHDSLQDKLPVADIKAVVTGKDCPHMKEKGALKQNKEVLELAFSILYDSNCQLN

FIAPDKYEYLIWTLGLESLLGKEKNEQELLELDKWASLWNWF

5-Helix- His6

MQLLSGIVQQQNNLLRAIEAQQHLLQLTVWGIKQLQARILAGGSGGHTTWMEWDREIN

NYTSLIHSLIEESQNQQEKNEQELLEGSSGGQLLSGIVQQQNNLLRAIEAQQHLLQLTVW

GIKQLQARILAGGSGGHTTWMEWDREINNYTSLIHSLIEESQNQQEKNEQELLEGSSGGQ

LLSGIVQQQNNLLRAIEAQQHLLQLTVWGIKQLQARILAGGHHHHHH

REFERENCES

1. Rivas, J. and C. Fontanillo, *Protein-Protein Interactions Essentials: Key Concepts to Building and Analyzing Interactome Networks.* PLoS Computational Biology, 2010. **6**(6): p. e1000807.
2. Walker, S.N., et al., *GLUE That Sticks to HIV: A Helix-Grafted GLUE Protein That Selectively Binds the HIV gp41 N-Terminal Helical Region.* ChemBioChem, 2015. **16**(2): p. 219-222.
3. Fosgerau, K. and T. Hoffmann, *Peptide therapeutics: current status and future directions.* Drug Discovery Today, 2015. **20**(1): p. 122-128.
4. Angelo, N.G. and P.S. Arora, *Nonpeptidic Foldamers from Amino Acids: Synthesis and Characterization of 1,3-Substituted Triazole Oligomers.* Journal of the American Chemical Society, 2005. **127**(49): p. 17134-17135.
5. Catherine, M.G., et al., *Foldamers as versatile frameworks for the design and evolution of function.* Nature Chemical Biology, 2007. **3**(5): p. 252.
6. Chapman, R.N., G. Dimartino, and P.S. Arora, *A highly stable short alpha-helix constrained by a main-chain hydrogen-bond surrogate.* Journal of the American Chemical Society, 2004. **126**(39): p. 12252.
7. Davis, J.M., L.K. Tsou, and A.D. Hamilton, *Synthetic non-peptide mimetics of α-helices.* Chemical Society Reviews, 2007. **36**(2): p. 326-334.
8. Gellman, S., *Foldamers: A manifesto*, in *Accounts Chem. Res.* 1998. p. 173-180.
9. Joshua, A.K., *Stapled peptides: Magic bullets in nature's arsenal.* Nature Chemical Biology, 2010. **6**(8): p. 566.

10. Ross, N.T., W.P. Katt, and A.D. Hamilton, *Synthetic mimetics of protein secondary structure domains.* Philosophical Transactions of the Royal Society A, 2010. **368**(1914): p. 989-1008.

11. Kritzer, J.A., et al., *Miniature Protein Inhibitors of the p53–hDM2 Interaction.* ChemBioChem, 2006. **7**(1): p. 29-31.

12. Lemmon, M.A., *Pleckstrin homology domains: not just for phosphoinositides.* Biochemical Society Transactions, 2004. **32**: p. 707-711.

13. Lemmon, M.A., *Pleckstrin homology (PH) domains and phosphoinositides.* Cell Biology Of Inositol Lipids And Phosphates, 2007. **74**: p. 81-93.

14. Lemmon, M.A. and K.M. Ferguson, *Pleckstrin homology domains.* Current topics in microbiology and immunology, 1998. **228**: p. 39.

15. Buzon, V., et al., *Crystal Structure of HIV-1 gp41 Including Both Fusion Peptide and Membrane Proximal External Regions (Structure of gp41 with MPER and FPPR).* PLoS Pathogens, 2010. **6**(5): p. e1000880.

16. Chan, D.C., et al., *Core Structure of gp41 from the HIV Envelope Glycoprotein.* Cell, 1997. **89**(2): p. 263-273.

17. Chan, D.C. and P.S. Kim, *HIV Entry and Its Inhibition.* Cell, 1998. **93**(5): p. 681-684.

18. Eckert, D.M. and P.S. Kim, *Design of Potent Inhibitors of HIV-1 Entry from the gp41 N-Peptide Region.* Proceedings of the National Academy of Sciences of the United States of America, 2001. **98**(20): p. 11187-11192.

19. Lalezari, P.J., et al., *A phase II clinical study of the long-term safety and antiviral activity of enfuvirtide-based antiretroviral therapy.* AIDS, 2003. **17**(5): p. 691-698.

20. Patel, I., et al., *Pharmacokinetics, Pharmacodynamics and Drug Interaction Potential of Enfuvirtide.* Clinical Pharmacokinetics, 2005. **44**(2): p. 175-186.

21. Root, M. and H. Steger, *HIV-1 gp41 as a Target for Viral Entry Inhibition.* Current Pharmaceutical Design, 2004. **10**(15): p. 1805-1825.

22. Komander, D., et al., *An InCytes from MBC Selection: An α-Helical Extension of the ELMO1 Pleckstrin Homology Domain Mediates Direct Interaction to DOCK180 and Is Critical in Rac Signaling.* Cell Molecular Biology, 2008. **19**(11): p. 4837-4851.

23. Horne, W.S., et al., *Structural and biological mimicry of protein surface recognition by α/β-peptide foldamers.* Proceedings of the National Academy of Sciences, 2009. **106**(35): p. 14751.

24. Root, M.J., M.S. Kay, and P.S. Kim, *Protein Design of an HIV-1 Entry Inhibitor.* Science, 2001. **291**(5505): p. 884-888.

25. Alain, G., et al., *A new reporter cell line to monitor HIV infection and drug susceptibility in vitro.* Proceedings of the National Academy of Sciences of the United States of America, 1997. **94**(9): p. 4653.

26. Haché, G., et al., *Evolution of HIV-1 Isolates that Use a Novel Vif-Independent Mechanism to Resist Restriction by Human APOBEC3G.* Current Biology, 2008. **18**(11): p. 819-824.

27. Tennyson, R.L., et al., *Helix-Grafted Pleckstrin Homology Domains Suppress HIV-1 Infection of CD4-Positive Cells.* ChemBioChem, 2016. **17**(20): p. 1945-1950.

CHAPTER FOUR

Evaluation of Sequence Variability in HIV-1 gp41 C-peptide Helix-Grafted Proteins[1]

In this work, the protein engineering design and initial analysis was done by both Rachel Tennyson and myself, as well as cloning and protein purification. I carried out the library design, execution, and analysis experiments. Terumasa Ikeda at the University of Minnesota analyzed all proteins in the live virus assays.

4.1 Introduction

The large size, diverse structures, and molecular complexity found in proteins often generates surfaces that potently and selectively bind to therapeutically-relevant targets that challenge—or evade—small-molecules, including protein-protein interactions (PPIs).[1] Additionally, the discovery of protein binders, through laboratory evolution, is often higher-throughput and simpler than the analogous small-molecule centered discovery process.[2] A number of therapeutically significant PPIs feature a cleft that is shape-complementary to an exposed α-helix. Given this, the inhibition of PPIs that feature a helix and helix-binding cleft at the interface, has been the focus of many therapeutic discovery efforts.[3-15]

[1] Reprinted with permission from:
Tennyson, R.L.*, Walker, S.N.*, Ikeda, T., Harris, R.S., McNaughton, B.R., Evaluation of Sequence Variability in HIV-1 gp41 C-peptide Helix-Grafted Proteins, *Bioorganic & Medicinal Chemistry*, 2017, 26, 1220
*Both authors contributed equally

In the previous chapters I described a new approach to folded helix display that relies on protein engineering.[16, 17] A stable scaffold protein displaying a helix, which contains at least one solvent-exposed face, serves as a generic foundation which can be designed to mimic, or extend, any desired helical interface. We term this method 'helix-grafted display'. In developing a general scaffold for helix-grafted display, we sought to identify a protein fold that presents a properly folded α-helix in such a way as to permit direct receptor access to one helical face. Ideally, the scaffold would be stable, express well as a soluble protein in *E. coli*, and tolerate multiple mutations and/or extension of the helix. Our initial scaffold search identified Pleckstrin Homology (PH) domains as suitable starting points. From this work, we identified ELMO, a PH domain amenable to helix-grafting (Figure 4.1A).[16]

We have used helix-grafted display to engineer proteins capable of suppressing HIV-1 entry.[16] This therapeutic activity is the result of inhibiting an intramolecular PPI between HIV-1 gp41 C-peptides (α-helix) and a trimer of HIV-1 gp41 N-peptides, which contain C-peptide α-helix-binding clefts. This PPI generates a 'trimer-of-hairpin' assembly that brings HIV and mammalian cell membranes into close proximity, leading to membrane fusion and HIV-1 entry.[18, 19] Helix-grafted proteins presenting a C-peptide surrogate bind pre-fusogenic gp41 N-peptide trimer, leading to suppression of gp41 fusogenic assembly.

Cpep-ELMO consists of the ELMO scaffold (with a truncated N-terminal helix) and a genetic fusion between the N-terminal ELMO helix and gp41 C-peptide (Figure 4.1B). Cpep-ELMO binds tightly to gp41 5-helix (K_D ~ 90 nM) and suppresses entry of HIV-1 into Cluster of Differentiation 4 (CD4) positive human cells.[16] In contrast to its peptide counterpart (the FDA approved peptide drug Fuzeon™, or Enfuvirtide), Cpep-ELMO is stable in human serum for greater than 12 hours.[20-22]

During the course of these earlier studies, we questioned whether or not the gp41 C-peptide helix could be evolved for improved function, within the context of helix-grafted display. We reasoned that mutation and optimization of solvent-exposed residues on the grafted C-peptide helix could result in improved N-peptide trimer recognition, improved expression, and/or improved stability of the helix-grafted protein. Additionally, we reasoned that mutating solvent exposed C-peptide residues, and evaluating this protein library for target affinity, would provide valuable information on the relative role each mutated C-peptide residue plays in gp41 N-peptide trimer recognition, and the capacity for variation within this helix. With this in mind, we set out to create a Cpep-ELMO library, with randomized solvent-exposed residues in the grafted helix, and use yeast display to evaluate this helix-grafted display protein library for N-peptide trimer recognition.

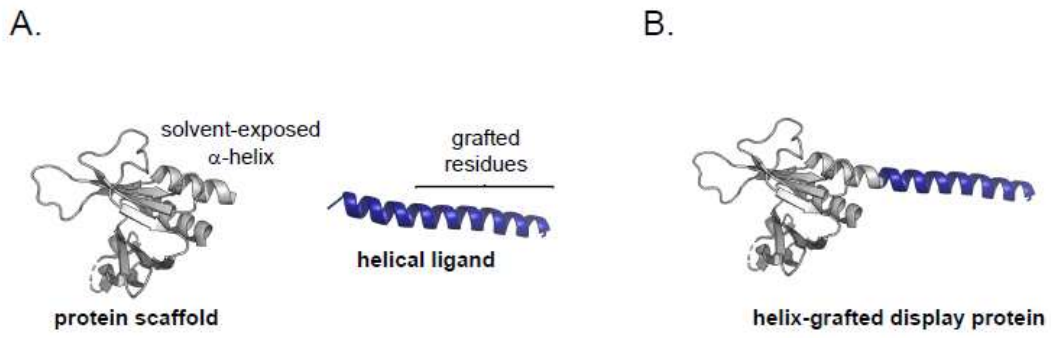


Figure 4.1 (A) Protein scaffold truncated ELMO (grey, PDB: 2VSZ) with N- and C-terminal solvent-exposed helices previously found to be an ideal candidate for grafting the helical ligand gp41 Cpeptide (density blue, PDB: 1AIK). (B) Helix-grafted display technology: a therapeutically-relevant helix (HIV gp41 C-peptide) is grafted onto N-terminal solvent exposed helix on ELMO, a Pleckstrin Homology (PH) domain to generate helix-grafted display protein Cpep-ELMO.

4.2 Yeast display evolution of a Cpep-ELMO helix-grafted display protein library

In the previous chapter, we had first demonstrated that we could graft a portion of gp41 C-peptide onto ELMO, to generate a new protein that reliably replicates C-peptide display and suppresses HIV-1 entry.[16] We next sought to evaluate sequence variability in the grafted helix.

Five positions on Cpep-ELMO (Figure 4.2A) that bury into the C-peptide helix-binding cleft of N-peptide trimer (wild-type sequence WWIYI) were randomized by saturation mutagenesis. In particular, the tryptophan residues (labeled residues 1 and 2 in Figure 4.2A) we selected to randomize are buried in a deep (and conserved) hydrophobic pocket[18]; their deletion has been reported to significantly reduce affinity[23]. Thus, these mutations, in a sense, serve as a control for amino acid evolution ('hits' should regain those tryptophan residues).

The ~3.2 million-member Cpep-ELMO derived protein library was displayed on yeast, and tightest binders were selected by cell sorting. Specifically, the yeast display library was made to express Cpep-ELMO derived library members that contain a C-terminal cMyc tag. Following expression of our displayed library, yeast was concomitantly incubated with a FITC-labelled anti-cMyc antibody to confirm expression and display of our proteins and biotinylated gp41 5-helix.[7, 24] 5-helix is a previously reported protein that covalently tethers five of the six subunits within the gp41 trimer-of-hairpins so that when folded, it features the gp41 coiled-coil with two of its binding sites already occupied, and just a single exposed interface. This reduces the recognition event to a simple 1:1 interaction, simplifying the analysis. Following a brief washing step, streptavidin-phycoerythrin conjugate (SA-PE) was incubated with the complex. Thus, yeast displaying biotin, by virtue of a surface protein/5-helix interaction, will be bound by SA-PE. Yeast were then analyzed by flow cytometry and sorted for the highest FITC fluorescence intensity (which comments on display efficiency) and PE red pigment intensity (which comments on 5-helix binding efficiency, summarized in Figure 4.2B). Yeast that display tightest affinity proteins were enriched by flow cytometry over the course of two rounds of screening against 500 nM or 100 nM biotinylated 5-helix, respectively (Figure 4.2C).

From our flow cytometry experiments, we identified eleven unique sequences that retain affinity for 5-helix in our yeast display screen. The established preference for tryptophan at positions W1 and W2 was faithfully replicated, suggesting that the method is sensitive to tightest receptor affinity. Meanwhile, while hydrophobic residues were often selected for, the other positions interrogated in this study are reasonably amenable to mutation, supporting the idea that variation within certain residues on the binding interface is permitted (Figure 4.2D). We next evaluated the effect helix mutations have on structure, expression, stability, function (5-helix recognition), and suppression of HIV-1 entry in a live virus assay.

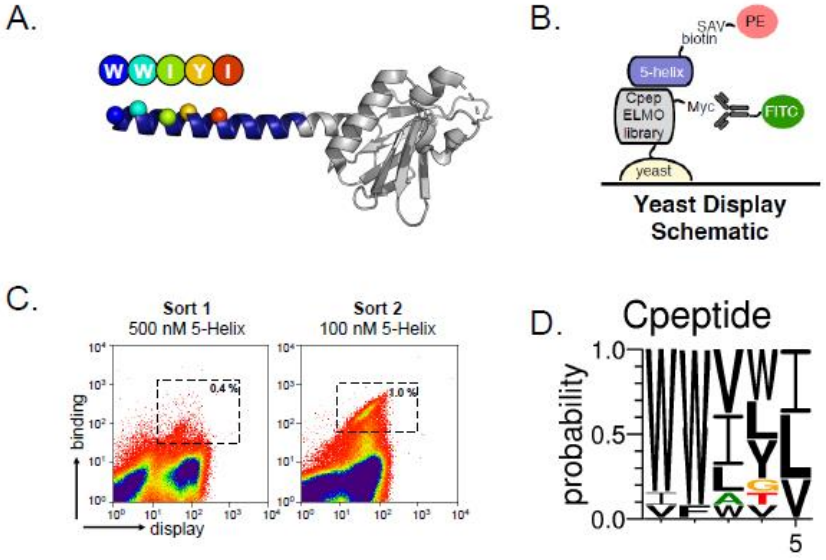


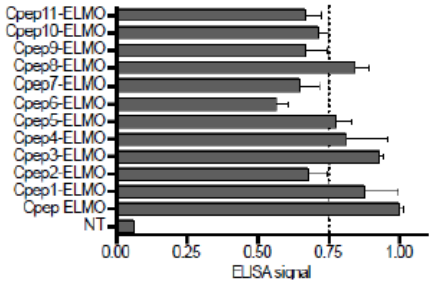
Figure 4.2 (A) Library design for Cpep-ELMO helix-grafted display library. Residues labeled with wildtype sequence, and highlighted as rainbow spheres, were mutated to all possible 20 proteinogenic amino acids, generating a ~3.2 million-member protein library. (B) Flow cytometry enrichment scheme to identify novel Cpep-ELMO helix-grafted proteins that bind to gp41 5-helix. (C) Flow cytometry data from yeast display library screening (sort 1: left; sort 2: right). Yeast selected in round 1 and round 2, representing 0.4% and 1% of the total population, respectively, are within dashed boxed. (D) Sequence logo generated from 11 unique Cpep-ELMO derived mutants enriched after two rounds of yeast display screening.

4.3 Evolved Cpep-ELMO mutants retain affinity for HIV-1 gp415-helix

Given the findings from our yeast display screening, whether or not solvent exposed residues on the gp41 C-peptide helix are amenable to mutation is no longer in question. What remained a question, however, was what effect these mutations had on the affinity for gp41 5-helix. We assessed binding between Cpep-ELMO mutants and gp41 5-helix by *in vitro* ELISA.

Biotinylated gp41 5-helix (Supplemental Figure 6.3A) was first immobilized onto commercially available streptavidin coated plates, which were then washed to remove unbound protein. Enriched mutants (Cpep1-ELMO – Cpep11-ELMO) containing a C-terminal FLAG tag were separately incubated with immobilized 5-helix and a commercially available anti-FLAG antibody conjugated to HRP. Binding was analyzed by measuring absorbance, following incubation with TMB One™ substrate. Of the initial eleven hits, five proteins (Cpep1, 3, 4, 5, 8-ELMO) compared favorably to Cpep-ELMO and warranted further investigation (Figure 4.3A). Amino acid composition of Cpep-ELMO, and our evolved proteins at mutated positions is shown in Figure 4.3B.

A.



B.

position	1	2	3	4	5
Cpep-ELMO	W	W	I	Y	I
Cpep1-ELMO	W	W	V	V	I
Cpep3-ELMO	W	W	V	Y	L
Cpep4-ELMO	W	W	I	T	V
Cpep5-ELMO	W	W	V	L	V
Cpep8-ELMO	W	W	L	Y	L

Figure 4.3 (A) ELISA data showing binding between immobilized gp41 5-helix and Cpep-ELMO, or mutants described in this work. Error bars represent the standard error of the mean from three replicate experiments. (B) Sequence of Cpep-ELMO, and mutants selected for by yeast display that compare favorably in ELISA experiment to Cpep-ELMO (Cpep1-ELMO, Cpep3-ELMO, Cpep4-ELMO, Cpep5-ELMO, and Cpep8-ELMO).

4.4 Evolved Cpep-ELMO proteins are structured, stable, and in some cases, express more robustly in *E. coli*

Given the findings from our yeast display screening and ELISA experiments, the solvent-exposed residues on the gp41 C-peptide helix were found to be amenable to mutation and retain function of binding the target gp41 helical bundle. What remained to answer, however, was what effect these mutations had on the larger protein structure. We began by comparing protein expression levels of mutants to our starting protein (Cpep-ELMO). In addition to selecting for improved target affinity, display-based methods often select for proteins with improved stability and expression[25]. We expressed five of the tightest binding Cpep-ELMO mutants in *E. coli* as N-terminal His6x fusions, and purified them by nickel-NTA chromatography. Levels of purified proteins were then measured as mg/L of culture with spectroscopy (nanodrop analysis). As shown in Figure 4.4A, all but one of our evolved proteins expressed at higher levels, compared to Cpep-ELMO. While Cpep-ELMO expressed modestly in *E. coli* (~2.5 mg/L), our best performing mutants, Cpep3-ELMO and Cpep4-ELMO, expressed more than two-fold higher (Figure 4.4A and Supplemental Figure 6.3B).

In addition to expressing better in *E. coli*, evolved mutants retain structural features associated with our starting protein (Figure 4.4B) as assessed by circular dichroism (CD). This makes sense: high efficiency yeast display requires proper protein folding and shuttling to the cell surface. Our yeast display screen concomitantly selects for robust expression, protein stability, and high affinity for gp41 N-peptide.

Having established that mutations selected for in our yeast display experiment do not appreciably alter protein structure, we next evaluated the effect yeast display evolution had on protein stability. Relative stabilities of Cpep-ELMO, and evolved mutants were measured by

thermally denaturing (0 °C – 90 °C) each protein and measuring helicity by CD at 222 nm (Figure 4.4C). Mutated proteins unfolded within the range 44–57°C, which is not dramatically lower than what is the observed melting temperature for Cpep-ELMO (57 °C). Collectively, these findings show that proteins evolved in our yeast display screen not only retained affinity for gp41 5-helix, but these proteins express—sometimes more robustly—as soluble proteins in *E. coli*, and these proteins retain their structure and stability.

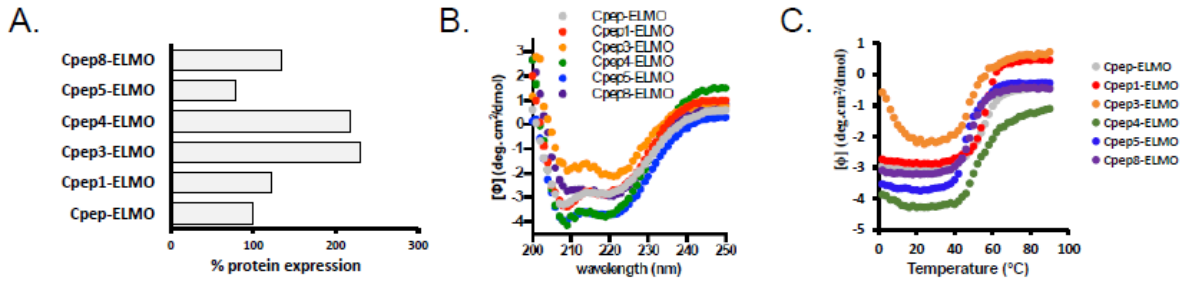


Figure 4.4 (A) Expression levels, in *E. coli*, for our starting helix-grafted display protein (Cpep-ELMO) and mutants generated from our yeast display screen (Cpep1-ELMO, Cpep3-ELMO, Cpep4-ELMO, Cpep5-ELMO, and Cpep8-ELMO). (B) Circular dichroism spectra for Cpep-ELISA and mutants described in this work showing most mutants retained similar secondary structural composition. (C) Melting data for Cpep-ELMO and mutants described in this work showing mutants retaining protein thermal stability.

4.5 Evolved Cpep-ELMO mutants suppress HIV infection in a live virus assay

Having demonstrated that some of our evolved C-peptide ELMO helix-grafted display proteins express better than Cpep-ELMO, retain helical structure, and bind to gp41 5-helix in vitro, we next measured their ability to suppress HIV infection using a live virus assay we have previously reported[16, 26]. In this method, HIV IIIB is administered to CD4+ mammalian cells stably integrated with a plasmid that encodes the HIV-1 long-terminal repeat (LTR) upstream of green fluorescent protein. Since HIV-1 Tat/TAR-dependent transcription ultimately leads to the expression of GFP, we will be able to see if HIV-1 infects these cells. The number of cells that express GFP correlates to the number of cells infected with the virus and can thus be measured via

flow cytometry. As seen in Figure 4.5, all but one of the evolved helix-grafted display proteins (Cpep8-ELMO) inhibit HIV-1 entry better than the original helix-grafted display protein, Cpep-ELMO. A number of our mutated proteins have dramatically improved Half-Maximal Inhibitory Concentration (IC_{50}) values for suppression of HIV-1 entry (IC_{50}: Cpep1-ELMO: 0.43 μM; Cpep3-ELMO: 0.31 μM; Cpep4-ELMO: 0.77 μM; Cpep5-ELMO: 0.19 μM. In contrast, the IC_{50} of our starting protein, Cpep-ELMO, was not reached over the course of the concentrations used in this experiment.

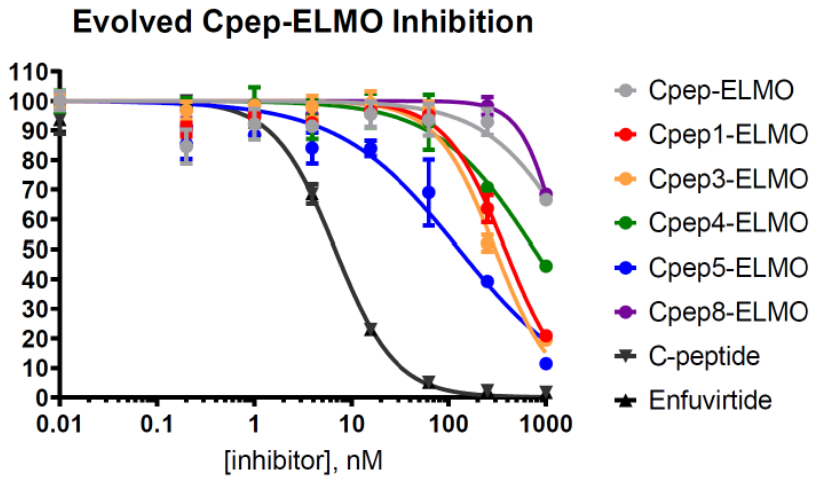


Figure 4.5 Normalized suppression of HIV-1 entry in CD4-positive cells in a live virus assay by Cpep-ELMO and derivatives. Each data point reports HIV-1 infectivity relative to the peptide hydrating buffer treated control (mean +/− SEM from 3 replicates; error bars not shown are smaller than the data points).

4.6 Conclusion

Many disease-relevant protein-protein interactions (PPIs) utilize interfaces that involve an α-helix and helix-binding groove. In order to disrupt these PPIs, researchers have developed various techniques to stabilize helical display of side chains within a peptide or peptide mimetic. While these techniques, such as peptide 'stapling', 'hydrogen-bond surrogate' engineering, and helical 'foldamers' represent important sectors of biomimetic research, and are important reagents

for inhibiting disease-relevant PPI's that feature a helix-binding groove, their synthesis and purification can be expensive and laborious. Researchers have also described 'minimal proteins' with an evolved helix (for tailored recognition); however, these polypeptides do not express as a folded protein in *E. coli* and must be prepared by solid-phase peptide synthesis, which is also laborious and expensive.

In the previous chapters I discussed 'helix-grafted display', a potentially general solution for displaying a folded therapeutically-relevant helix on a protein scaffold. In Chapters 2 and 3, we showed that a Pleckstrin Homology (PH) domain can be resourced as a helix-grafted display scaffold, wherein a solvent-exposed helix serves in the wild-type protein serves as a generic canvas upon which to paint, or extend, any desired helical interface. My research has focused on the development of new proteins that display HIV-1 gp41 C-peptide, an established therapeutic reagent. We showed that Cpep-ELMO, an engineered protein that displays gp41 C-peptide as a fusion to the N-terminal helix of the ELMO PH domain, binds to gp41 5-helix, which mimics the pre-fusogenic state of gp41 and serves as a therapeutically-relevant target. Cpep-ELMO potently binds to gp41 5-helix (K_D ~ 90 nM) and suppresses HIV entry into CD4-positive cells, in a concentration-dependent manner.

In this work, we sought to optimize properties of Cpep-ELMO mutants by mutating five C-peptide binding face residues to all possible proteinogenic amino acids and selecting for efficient display on yeast (which relates to stability) and 5-helix recognition. As a result of this screen, we reasoned that we could not only improve recognition, but also assess the variability of sequences within the solvent-exposed binding face of C-peptide in our helix-grafted protein. After screening a ~3.2 million-member Cpep-ELMO derived library, we found that two neighboring tryptophan residues (labeled W1 and W2 in this work, Figure 4.2A) are enriched for, supporting previous

findings detailing their importance in gp41 5-helix recognition. Thus, their enrichment demonstrated the ability of our yeast display screen to enrich for residues necessary for 5-helix recognition. Other residues, however, (labeled I3, Y4, and I5 in this work, Figure 4.2A) exhibit sequence variability. While hydrophobic residues are generally selected for at these positions, no single residue emerged as absolutely necessary, or even particularly favored. Using eleven proteins (Cpep1-ELMO–Cpep11-ELMO) enriched in our screen, we measured their expression in *E. coli*, structure, and stability. We also measured the ability of these proteins to recognize gp41 5-helix and suppress HIV entry using a live virus assay. These characteristics and activities were compared to our starting protein, Cpep-ELMO.

Satisfyingly, 5 of the 11 favor comparably to Cpep-ELMO when it comes to 5-helix recognition; all 5 of these proteins exhibit CD spectra that indicates retention of helical structure; all express as soluble proteins in *E. coli*, and in some cases, expression levels are improved, in comparison to the starting protein. Finally, 4 of our evolved proteins (Cpep1-ELMO; Cpep3-ELMO; Cpep4-ELMO; Cpep5-ELMO) suppress HIV-1 entry better than our starting helix-grafted protein (Cpep-ELMO). This work shows that solvent-exposed residues on C-peptide are amendable to mutation, and that their mutation can generate new proteins with improved properties and therapeutic activity.[27]

4.7 Methods

Protein Expression and Purification

Genes were cloned into pET using restriction enzymes BamHI and KpnI, downstream of a His6x tag and transformed into BL21s (DE3). Cells were grown in 1 L 2XYT cultures containing 100 μg/mL carbenicillin at 37 °C to OD_{600} =0.5 - 0.8 and induced with 1 μM IPTG at 25 °C overnight.

Cells were then collected by centrifugation, resuspended in PBS buffer (20 mM Na_2HPO_4 pH 7.4, 100 mM NaCl) and stored at -20 °C. Frozen pellets were thawed and sonicated with 1 second pulses for 3 minutes. The lysate was cleared by centrifugation (8,000 rpm 10 min.) and the supernatant was mixed with 1 mL of Ni-NTA agarose resin for 30 min at 4 °C. The resin was collected by centrifugation (4300 rpm, 10 min.). The resin was sequentially washed with 30 mL of PBS buffer containing 20 mM imidazole, 50 mL buffer containing 50 mM imidazole, and 15 mL buffer containing 75 mM imidazole. The protein was then eluted with 4 mL PBS buffer containing 400 mM imidazole. The proteins were dialyzed against PBS buffer and analyzed for purity by SDS-PAGE. Purified protein concentrations were quantified using Beer's Law at an absorbance of 280 nm, following standard practice.1 Protein expression was calculated as a measure of eluted protein yield in mg/L of induced *E. coli* culture.

Resolubilization of 5helix Inclusion Bodies

5helix-His6x was cloned into a modified pETDuet-1 vector using restriction enzymes NdeI and KpnI and transformed into BL21s (DE3). Cells were induced to express 5helix-His6x and lysed as described above. The lysate was cleared by centrifugation (8,000 rpm, 20 min.) and the supernatant discarded. The pellet was washed twice with PBS buffer containing 0.5 % Triton® X-100 and once with PBS buffer. The pellet was resuspended in urea buffer (PBS buffer with 8 M urea and 10 mM imidazole) to resolubilize the inclusion bodies and cleared by centrifugation (8,000 rpm, 30 min.) The supernatant was mixed with 1 mL of Ni-NTA agarose resin for 1 hour at 4 °C. The resin was collected by centrifugation (4,300 rpm, 4 min.). The resin was washed with 50 mL of urea buffer and eluted with 40 mL of urea elution buffer (PBS buffer with 6 M urea and 100 mM imidazole) into 460 mL PBS buffer by gravity elution while stirring to refold the protein. The 500 mL elution was run through a column containing 1 mL of Ni-NTA agarose resin and eluted with

5 mL PBS buffer containing 400 mM imidazole. The protein was dialyzed against PBS buffer and analyzed for purity by SDS-PAGE. Refolding analysis was conducted by CD. Purified proteins were quantified using Beer's Law at an absorbance of 280 nm. Protein expression was calculated as a measure of eluted protein yield in mg/L of induced *E. coli* culture.

Circular Dichroism

Proteins were purified as described above. Separately, each protein was diluted to 3-10 μM in PBS buffer. Wavelength data are the average of three scans from 250 nm to 200 nm in 1 nm steps at 25 °C. Thermal denaturation experiments at 222 nm were run from 0 to 90 °C in two-degree steps at a two-degree/minute rate of increase with one-minute equilibration and data averaging at each temperature. Tm values were obtained from minima of the first derivative of θ versus 1/T plots.

Biotinylation

5helix-His6x was cloned into a pET vector containing an upstream Avitag™ (GLNDIFEAQKIEWHE)-GGSGGSGGT linker using restriction enzymes KpnI and PacI. The protein was resolublized from inclusion bodies as described above in PBS buffer. His6x-BirA was cloned into pET using restriction enzymes NcoI and KpnI and purified as described above in Phosphate buffer. 300 μL of Avitagged™ 5-Helix-His6 protein at 38 μM was incubated with 6 μL of His6-BirA at 1 mg/mL using Avidity® BirA biotin-protein ligase standard reaction kit at 30°C for 40 min. Biotinylation was confirmed by Agilent 6220 TOF LC-MS.

***In vitro* ELISA**

Helix-grafted Cpep-ELMO proteins were cloned into a pET vector upstream of a C-terminal His6x-GS-FLAG tag (DYKDDDDK) using restriction enzymes NheI and BamHI. Proteins were purified as described above in PBS buffer. The biotinylated AvitaggedTM 5-helix-His6 (biotin-5helix) was prepared as described above in PBS buffer and diluted to 10 μg/mL. Pierce®

Streptavidin coated clear 96-well plates with a binding capacity of 5 pmol were pre-blocked with 200 μL of wash buffer (PBS buffer, 0.5 mg/mL BSA, 0.1% tween® 20) for 1 hour. Biotin-5helix was immobilized on the streptavidin-coated plates by incubating 100 μL of diluted protein for 1 hour at room temperature, followed by 4x 200 μL washes (5 min). 100 μL of Cpep-ELMO proteins were incubated at a concentration of 75 nM for 1 hour, followed by 4x 200 μL 4°C washes (5 min). 100 μL of a 1:10,000 dilution of HRP-conjugated mouse anti-DDDDK antibody in Odyssey® Blocking Buffer was incubated for 1 hour at room temperature, followed by 4x 200 μL 4°C washes (5 min). Color was developed using TMB-One substrate and absorbance was measured at 655 nm on a SynergyMx Microplate Reader.

Infectivity Assay

All proteins were shipped on ice in a 20% glycerol-PBS stock. HEK293T cells were maintained in DMEM supplemented with 10% fetal bovine serum (FBS) and 0.5% penicillin/streptomycin (P/S). CEM-GFP cells were cultured in RPMI with 10% FBS and 0.5% P/S. The procedure was virtually identical to a previously published variant. The HIV-1 IIIB C200 proviral expression construct has been described previously. Viruses were produced by transfection of 3.0 μg of Vif-proficient proviral expression construct into 293T cells ($3.0x10^6$) using TransIT®-LT1 reagent. 48 hr later, virus-containing supernatants were filtered by 0.45 μm filters and used to infect into $2.5x10^4$ CEM-GFP cells with varying concentration of inhibitors. Infectivity (GFP+ cells) was measured by flow cytometry at 2 days post-infection.

Protein Library Preparation

EBY100 yeast (trp-, leu-, with the Aga1p gene stably integrated) and the pCTCON2 plasmid were generously provided by the Wittrup lab (MIT). The gene encoding Cpep-ELMO was amplified by PCR and cloned into pCTCON2 in-frame with Aga2, an N-terminal HA-tag, and a C-terminal myc

tag, using the restriction enzymes NheI and BamHI. When analyzed for display only, Cpep-ELMO displayed efficiently on EBY100 cells (~60-85 %, data not shown). Next, the Cpep-ELMO library was created by amplifying the Cpep-ELMO gene with 5 sites in the N-terminal α-helical region (Trp1, Trp4, Ile8, Tyr11, and Ile15) substituted with NNK codons. Using restriction enzymes NheI and BamHI, the library amplicon was digested and inserted into a digested pCTCON2 plasmid previously containing an insert with several successive stop codons to prevent false positive screening (nonsense pCTCON2). The resulting DNA plasmid was used as a template for a second PCR with homologous recombination primers, Fwd: (5′-CTC TGG TGG AGG GCG TAG CGG AGG CGG AGG GTC GGC TAG C-3′) and Rev: (5′-CGA GCT ATT ACA AGT CCT CTT CAG AAA TCA GCT TTT GTT CGG ATC C-3′), which are designed to create an insert with ~40 base pairs of overlap with the pCTCON2 vector. The resulting amplicon, containing the randomized sequences, was then cloned into pCTcon2 using homologous recombination in EBY100 yeast. Approximately 1 µg of nonsense pCTCON2 vector cut with BamHI and NheI was mixed with ~3 µg of the amplified library, ethanol precipitated, and transformed via electroporation into 50 µL of electrocompetent EBY100 using 2 mm cuvettes. 10 of these electroporations were performed and each was immediately rescued with 2 mL pre-warmed YPD and combined for 2 hours at 30 °C. After rescue, yeast were centrifuged at 2,500 xg for 1 minute, and supernatant YPD was removed. Yeast were resuspended in 1 mL fresh SD-CAA (5.4 g/L Na_2HPO_4, 8.6 g/L $NaH_2PO_4 \cdot H_2O$, 20 g/L dextrose, 6.7 g/L yeast nitrogen base lacking amino acids, 5 g/L casamino acids, 200 kU/L penicillin, 0.1 g/L streptomycin). A small portion was plated by serial dilution onto SD-CAA plates, and incubated at 30 °C for 3 days in order to determine the transformation efficiency. The remainder were cultured in 50 mL SD-CAA for yeast display screening. Efficiency of homologous

recombination transformation was determined to be 1 x 10^7 – 1.63 x 10^7 using the equation below. transformation efficiency = (# colonies x Vol quenched)/(dilution x Vol plated).

Yeast Display Screening

After 2-3 days of growth in SD-CAA, the library was sub-cultured in SD-CAA at an initial density of 0.5 x 10^7 cells/mL and grown to a density of 2.0 x 10^7 cells/mL. Yeast were subsequently sub-cultured in SG-CAA (Galactose containing induction media) to a concentration of 1.0 x 10^7 cells/mL and grown for 1-2 days shaking at 250 RPM at a temperature of 25 °C. For each round of screening, approximately 10^8 cells were pelleted and washed with 1 mL of 4 °C PBS-BSA (Corning CellGro PBS 1x with 1 g/L BSA filter sterilized). Yeast were subsequently incubated with biotin-5helix at the concentrations given in the table below and a 1:250 dilution of FITC-conjugated anti-myc antibody at room temperature. After incubation, the yeast cells were incubated on ice for 5 minutes, pelleted at 12,000 xg for 30 seconds 4 °C and washed with 1 mL ice-cold PBS-BSA. The yeast was pelleted again and incubated with a 1:100 dilution of SA-PE in PBS-BSA on ice for 1 hour. After incubation, a final wash with ice-cold PBS-BSA was performed, and yeast positive for both FITC (display) and R-Phycoerythrin (5helix-binding) were sorted into 7 mL of SD-CAA media using a MoFlo Flow Cytometer (Beckman-Coulter). Sorted yeast were transferred to 50 mL of pre-warmed SD-CAA and incubated at 30 °C for 3 days shaking at 250 RPM. Additionally, plasmid DNA was recovered from the sorted library using a Zymoprep yeast plasmid miniprep II kit. This DNA was used to transform Invitrogen, Top10 *E. coli*.

4.8 Sequences of Proteins Used in this Work

5helix-His6

MQLLSGIVQQQNNLLRAIEAQQHLLQLTVWGIKQLQARILAGGSGGHTTWMEWDREIN
NYTSLIHSLIEESQNQQEKNEQELLEGSSGGQLLSGIVQQQNNLLRAIEAQQHLLQLTVW
GIKQLQARILAGGSGGHTTWMEWDREINNYTSLIHSLIEESQNQQEKNEQELLEGSSGGQ
LLSGIVQQQNNLLRAIEAQQHLLQLTVWGIKQLQARILAGGHHHHHH*

Avi-5helix-His6

MGLNDIFEAQKIEWHEGGSGGSGGTQLLSGIVQQQNNLLRAIEAQQHLLQLTVWGIKQL
QARILAGGSGGHTTWMEWDREINNYTSLIHSLIEESQNQQEKNEQELLEGSSGGQLLSGI
VQQQNNLLRAIEAQQHLLQLTVWGIKQLQARILAGGSGGHTTWMEWDREINNYTSLIH
SLIEESQNQQEKNEQELLEGSSGGQLLSGIVQQQNNLLRAIEAQQHLLQLTVWGIKQLQA
RILAGGHHHHHH*

His6-ELMO (with a truncated C-terminal helix)

MGSSHHHHHHSQDPKQQRLNRLVEGTCFRKLNARRRQDKFWYCRLSPNHKVLHYGDL
EESPQGEVPHDSLQDKLPVADIKAVVTGKDCPHMKEKGALKQNKEVLELAFSILYDSNC
QLNFIAPDKHEYCIWTDGLNALLGK*

His6-Cpep-ELMO

MGSSHHHHHHSQDPWMEWDREINNYTSLIHSLIEESQNQQKQQRLNRLVEGTCFRKLN
ARRRQDKFWYCRLSPNHKVLHYGDLEESPQGEVPHDSLQDKLPVADIKAVVTGKDCPH
MKEKGALKQNKEVLELAFSILYDSNCQLNFIAPDKHEYCIWTDGLNALLGK*

His6-Cpep1-ELMO

MGSSHHHHHHSQDPWMEWDREVNNVTSLIHSLIEESQNQQKQQRLNRLVEGTCFRKLN

ARRRQDKFWYCRLSPNHKVLHYGDLEESPQGEVPHDSLQDKLPVADIKAVVTGKDCPH

MKEKGALKQNKEVLELAFSILYDSNCQLNFIAPDKHEYCIWTDGLNALLGK*

His6-Cpep2-ELMO

MGSSHHHHHHSQDPWMEWDRELNNWTSLIHSLIEESQNQQKQQRLNRLVEGTCFRKL

NARRRQDKFWYCRLSPNHKVLHYGDLEESPQGEVPHDSLQDKLPVADIKAVVTGKDCP

HMKEKGALKQNKEVLELAFSILYDSNCQLNFIAPDKHEYCIWTDGLNALLGK*

His6-Cpep3-ELMO

MGSSHHHHHHSQDPWMEWDREVNNYTSLLHSLIEESQNQQKQQRLNRLVEGTCFRKL

NARRRQDKFWYCRLSPNHKVLHYGDLEESPQGEVPHDSLQDKLPVADIKAVVTGKDCP

HMKEKGALKQNKEVLELAFSILYDSNCQLNFIAPDKHEYCIWTDGLNALLGK*

His6-Cpep4-ELMO

MGSSHHHHHHSQDPWMEWDREINNTTSLVHSLIEESQNQQKQQRLNRLVEGTCFRKLN

ARRRQDKFWYCRLSPNHKVLHYGDLEESPQGEVPHDSLQDKLPVADIKAVVTGKDCPH

MKEKGALKQNKEVLELAFSILYDSNCQLNFIAPDKHEYCIWTDGLNALLGK*

His6-Cpep5-ELMO

MGSSHHHHHHSQDPWMEWDREVNNLTSLVHSLIEESQNQQKQQRLNRLVEGTCFRKL

NARRRQDKFWYCRLSPNHKVLHYGDLEESPQGEVPHDSLQDKLPVADIKAVVTGKDCP

HMKEKGALKQNKEVLELAFSILYDSNCQLNFIAPDKHEYCIWTDGLNALLGK*

His6-Cpep6-ELMO

MGSSHHHHHHSQDPIMEWDREVNNWTSLLHSLIEESQNQQKQQRLNRLVEGTCFRKLN

ARRRQDKFWYCRLSPNHKVLHYGDLEESPQGEVPHDSLQDKLPVADIKAVVTGKDCPH

MKEKGALKQNKEVLELAFSILYDSNCQLNFIAPDKHEYCIWTDGLNALLGK*

His6-Cpep7-ELMO

MGSSHHHHHHSQDPWMEFDREANNLTSLIHSLIEESQNQQKQQRLNRLVEGTCFRKLN

ARRRQDKFWYCRLSPNHKVLHYGDLEESPQGEVPHDSLQDKLPVADIKAVVTGKDCPH

MKEKGALKQNKEVLELAFSILYDSNCQLNFIAPDKHEYCIWTDGLNALLGK*

His6-Cpep8-ELMO

MGSSHHHHHHSQDPWMEWDRELNNYTSLLHSLIEESQNQQKQQRLNRLVEGTCFRKL

NARRRQDKFWYCRLSPNHKVLHYGDLEESPQGEVPHDSLQDKLPVADIKAVVTGKDCP

HMKEKGALKQNKEVLELAFSILYDSNCQLNFIAPDKHEYCIWTDGLNALLGK*

His6-Cpep9-ELMO

MGSSHHHHHHSQDPWMEWDREINNWTSLLHSLIEESQNQQKQQRLNRLVEGTCFRKL

NARRRQDKFWYCRLSPNHKVLHYGDLEESPQGEVPHDSLQDKLPVADIKAVVTGKDCP

HMKEKGALKQNKEVLELAFSILYDSNCQLNFIAPDKHEYCIWTDGLNALLGK*

His6-Cpep10-ELMO

MGSSHHHHHHSQDPWMEWDREINNYTSLIHSLIEESQNQQKQQRLNRLVEGTCFRKLN

ARRRQDKFWYCRLSPNHKVLHYGDLEESPQGEVPHDSLQDKLPVADIKAVVTGKDCPH

MKEKGALKQNKEVLELAFSILYDSNCQLNFIAPDKHEYCIWTDGLNALLGK*

His6-Cpep11-ELMO

MGSSHHHHHHSQDPVMEWDREVNNWTSLIHSLIEESQNQQKQQRLNRLVEGTCFRKLN

ARRRQDKFWYCRLSPNHKVLHYGDLEESPQGEVPHDSLQDKLPVADIKAVVTGKDCPH

MKEKGALKQNKEVLELAFSILYDSNCQLNFIAPDKHEYCIWTDGLNALLGK*

His6-BirA

MKDNTVPLKLIALLANGEFHSGEQLGETLGMSRAAINKHIQTLRDWGVDVFTVPGKGY

SLPEPIQLLNAKQILGQLDGGSVAVLPVIDSTNQYLLDRIGELKSGDACIAEYQQAGRGR

RGRKWFSPFGANLYLSMFWRLEQGPAAAIGLSLVIGIVMAEVLRKLGADKVRVKWPND

LYLQDRKLAGILVELTGKTGDAAQIVIGAGINMAMRRVEESVVNQGWITLQEAGINLDR

NTLAAMLIRELRAALELFEQEGLAPYLSRWEKLDNFINRPVKLIIGDKEIFGISRGIDKQG

ALLLEQDGIIKPWMGGEISLRSAEKGSHHHHHH*

REFERENCES

1. Chapman, Alex M. and Brian R. McNaughton, *Scratching the Surface: Resurfacing Proteins to Endow New Properties and Function.* Cell Chemical Biology, 2016. **23**(5): p. 543-553.
2. Michael, S.P. and R.L. David, *Methods for the directed evolution of proteins.* Nature Reviews Genetics, 2015. **16**(7): p. 379.
3. Checco, J.W. and S.H. Gellman, *Targeting recognition surfaces on natural proteins with peptidic foldamers.* Current Opinion in Structural Biology, 2016. **39**: p. 96-105.
4. Henchey, L., A.L. Jochim, and P.S. Arora, *Contemporary strategies for the stabilization of peptides in the alpha-helical conformation.* Current Opinion In Chemical Biology, 2008. **12**(6): p. 692-697.
5. Hill, D., et al., *A field guide to foldamers.* Chemical Reviews, 2001. **101**(12): p. 3893-4011.
6. Horne, W. and S. Gellman, *Foldamers with Heterogeneous Backbones.* Accounts Of Chemical Research, 2008. **41**(10): p. 1399-1408.
7. Horne, W.S., et al., *Structural and biological mimicry of protein surface recognition by α/β-peptide foldamers.* Proceedings of the National Academy of Sciences, 2009. **106**(35): p. 14751.
8. Jonathan, W.N., R.M. Fesinmeyer, and H.A. Niels, *Designing a 20-residue protein.* Nature Structural Biology, 2002. **9**(6): p. 425.
9. Joshua, A.K., *Stapled peptides: Magic bullets in nature's arsenal.* Nature Chemical Biology, 2010. **6**(8): p. 566.

10. Kritzer, J.A., et al., *Miniature Protein Inhibitors of the p53–hDM2 Interaction.* ChemBioChem, 2006. **7**(1): p. 29-31.

11. Patigiri, A., A.L. Joichim, and P.S. Arora, *A Hydrogen Bond Surrogate Approach for Stabilization of Short Peptide Sequences in alpha-Helical Conformation.* Accounts Of Chemical Research, 2008. **41**(10): p. 1289-1300.

12. Smith, B.A., et al., *Minimally cationic cell-permeable miniature proteins via alpha-helical arginine display.* Journal of the American Chemical Society, 2008. **130**(10): p. 2948.

13. Verdine, G. and G. Hilinski, *Stapled Peptides for Intracellular Drug Targets.* Methods In Enzymology: Protein Engineering For Therapeutics, Vol 203, Pt B, 2012. **503**(1): p. 3-33.

14. Walensky, L.D. and G.H. Bird, *Hydrocarbon-stapled peptides: principles, practice, and progress.* Journal of medicinal chemistry, 2014. **57**(15): p. 6275.

15. Yin, H. and A.D. Hamilton, *Strategies for Targeting Protein–Protein Interactions With Synthetic Agents*. 2005: Weinheim. p. 4130-4163.

16. Tennyson, R.L., et al., *Helix-Grafted Pleckstrin Homology Domains Suppress HIV-1 Infection of CD4-Positive Cells.* ChemBioChem, 2016. **17**(20): p. 1945-1950.

17. Walker, S.N., et al., *GLUE That Sticks to HIV: A Helix-Grafted GLUE Protein That Selectively Binds the HIV gp41 N-Terminal Helical Region.* ChemBioChem, 2015. **16**(2): p. 219-222.

18. Chan, D.C., et al., *Core structure of gp41 from the HIV envelope glycoprotein.* Cell, 1997. **89**(2): p. 263-73.

19. Chan, D.C. and P.S. Kim, *HIV entry and its inhibition.* Cell, 1998. **93**(5): p. 681-4.

20. Abel, S., D. Back, and M. Vourvahis, *Maraviroc: pharmacokinetics and drug interactions.* Antiviral Therapy, 2009. **14**(5): p. 607-618.

21. Jenssen, H. and S.I. Aspmo, *Serum stability of peptides.* Methods in Molecular Biology, 2008. **494**: p. 177-186.

22. Patel, I., et al., *Pharmacokinetics, Pharmacodynamics and Drug Interaction Potential of Enfuvirtide.* Clinical Pharmacokinetics, 2005. **44**(2): p. 175-186.

23. David, C.C., T.C. Christine, and S.K. Peter, *Evidence that a prominent cavity in the coiled coil of HIV type 1 gp41 is an attractive drug target.* Proceedings of the National Academy of Sciences of the United States of America, 1998. **95**(26): p. 15613.

24. Root, M.J., M.S. Kay, and P.S. Kim, *Protein Design of an HIV-1 Entry Inhibitor.* Science, 2001. **291**(5505): p. 884-888.

25. Boder, E. and K. Wittrup, *Yeast surface display for directed evolution of protein expression, affinity, and stability*, in *Methods Enzymol.* 2000. p. 430-444.

26. Haché, G., et al., *Evolution of HIV-1 Isolates that Use a Novel Vif-Independent Mechanism to Resist Restriction by Human APOBEC3G.* Current Biology, 2008. **18**(11): p. 819-824.

27. Tennyson, R.L., et al., *Evaluation of sequence variability in HIV-1 gp41 C-peptide helix-grafted proteins.* Bioorganic & Medicinal Chemistry, 2018. **26**(6): p. 1220-1224.

CHAPTER FIVE

Evolved Biologics Designed for Monthly Administration Potently Inhibit Entry of Enfuvirtide-Resistant HIV-1[1]

In this work, the protein engineering design and initial analysis was done by both Rachel Tennyson and myself, as well as cloning and protein purification. I carried out the library design, execution, and analysis experiments. Terumasa Ikeda at the University of Minnesota analyzed all proteins in the live virus assays.

5.1 Introduction

Individuals at high risk for HIV-1 infection often take pre-exposure prophylaxis (PrEP), a cocktail consisting of Emtricitabine and Tenofovir (both reverse transcriptase inhibitors).[1] For patients already infected with HIV-1, therapeutics are available that act on a number of proteins necessary for HIV-1 proliferation. For example, the principal components of combination antiretroviral therapy (cART) are typically inhibitors of viral protease and reverse transcriptase enzymes.[2] Both PrEP and cART require daily administration due to the relatively short *in vivo* lifetime of these small-molecule drugs.[1, 2] Given the stigma associated with HIV infection, there is a significant psychological barrier to obtaining, storing, and taking—on a daily basis—HIV antiretroviral prophylaxis and therapeutics. Since PrEP and cART do not inhibit HIV-1 entry, even

[1] Reprinted with permission from:
Ikeda, T.*, Tennyson, R.L.*, Walker, S.N.*, Harris, R.S., McNaughton, B.R., Evolved Biologics Designed for Monthly Administration Potently Inhibit Entry of Enfuvirtide-Resistant HIV-1, *ACS Infectious Disease*, 2019.
*All authors contributed equally

miniscule levels of replication-competent viral genome integration by intracellular HIV-1 can lead to the formation of latent viral reservoirs, which make a cure virtually impossible. Since inhibition of virus entry, by definition, stops viral genome integration, entry inhibition may be preferred to protect against HIV-1 infection and the formation of latent viral reservoirs.

HIV infection is initiated by binding between HIV gp120 and the cluster of differentiation 4 (CD4) receptor on immune cells, such as T helper cells, monocytes, macrophages, and dendritic cells.[3, 4] This recognition event is followed by a conformational change in HIV gp120, which exposes another HIV protein – viral glycoprotein gp41. The C-terminus of gp41 inserts into the HIV membrane, while the N-terminus inserts into the membrane of the engaged immune cell. Physical entry of the viral contents into the bound immune cell requires conformational collapse of prefusogenic N-peptide and C-peptide trimers into a "trimer-of-hairpins" assembly that promotes membrane fusion – a mandatory step for infection.[5]

Only one FDA approved HIV entry inhibitor (Enfuvirtide) targets gp41.[6, 7] Enfuvirtide is a 36-residue synthetic peptide that binds pre-fusogenic gp41 N-peptide trimer, inhibiting formation of the "trimer-of-hairpins" assembly required for viral membrane fusion and HIV infection of a host immune cell. Enfuvirtide is costly to synthesize, has poor solubility and bioavailability, is sensitive to serum and interstitial proteases (in vivo half-life is approximately 4 hours) and thus requires daily parenteral administration.[8] Long-lasting Enfuvirtide conjugates have been reported[9]; however, they require additional chemical synthesis and purification. Since Enfuvirtide is a single peptide, mutation of HIV gp41 can lead to Enfuvirtide-resistance.[10, 11] Researchers have reported D-peptides[12], mixed α/β peptides[13], and longer helical scaffolds that display C-peptide residues.[14] However, all of these require laborious chemical synthesis and purification.

We have reported a technique called 'helix-grafted display': a protein engineering-based solution to generate biologics that target therapeutically-relevant helix-binding clefts. While likely applicable to the myriad of disease-relevant protein-protein interactions that involve a helix and complementary surface, our studies have focused on targeting the prefusogenic HIV-1 gp41 N-peptide trimer.[15, 16] We have focused on Pleckstrin Homology domains (PH domains) as a helix-grafted display scaffold, since these proteins are generally stable, express in *E. coli* (a common bioproduction host), and contain a solvent-exposed helix.[17-19] First, specific helix residues are mutated to mimic gp41 C-peptide helix. Once the native helix is utilized, the remaining gp41 C-peptide helix is then genetically fused. The resulting protein mimics a folded gp41-C-peptide helix. We have shown that GLUE-Cpep[16], and a second-generation helix-grafted protein, Cpep-ELMO (C26-ELMO)[15], bind to gp41 '5-helix', which researchers, including our lab, have used as a surrogate for pre-fusogenic N-peptide trimer.[20] 5-helix is a single polypeptide that contains three copies of gp41 N-peptide and two copies of C-peptide. This architecture presents one vacant C-peptide binding site, thus simplifying the interaction to a 1:1 complex between 5-helix and C-peptide (or a C-peptide helix-grafted display protein). While Cpep-ELMO binds to 5-helix, it exhibits modest inhibition of HIV entry (IC_{50} >1 μM). Evolving Cpep-ELMO through saturation mutagenesis and yeast display screening has improved the effective inhibition to an IC_{50} of ~190 nM (Cpep5-ELMO), but the limitations of an N-terminal graft decrease serum stability compared to C-terminal graft GLUE-Cpep and limit helical elongation.[21] On the basis of these findings, we set out to optimize the scaffold and evolve the solvent-exposed helix residues in order to identify a cocktail of potent HIV-1 entry inhibitors.

5.2 Helix-grafted Sac7d (Sac7d-Cpep) potently binds 5-helix in *E. coli* cell lysate.

Previous studies suggested a relationship between the size of the helix-grafted display scaffold and HIV-1 entry inhibition: smaller scaffolds are more effective.[15] We therefore set our sights on a miniature (<10 kDa) PH domain (or structurally related protein) that expresses well in *E. coli* and is amenable to the extensive helix mutagenesis required for grafting and sequence optimization. Analysis of the Protein Data Bank (PDB), and literature, revealed Sac7d – a 7 kDa protein originating from the hyperthermophilic archaeon *Sulfolobus* (PDB: 2XIW).[22] We engineered this protein using helix-grafted display to function as a C-peptide mimic and analyzed the minimalist protein for structural and functional fidelity. Helix-grafted Sac7d (Figure 5.2A) expresses as a soluble protein in *E. coli* (Supplemental Figure 6.4A), and circular dichroism experiments show that it maintains structural features found in the wild-type protein with extended helical character (Figure 5.2B).

When 5-helix-His6 is expressed in *E. coli* in the absence of a binding partner, it expresses as an inclusion body and must be refolded. This is likely due to aggregation cause by the solvent-exposed hydrophobic C-peptide binding site. However, when 5-helix-His6 is appropriately bound by another protein, effectively filling the hydrophobic C-peptide binding site, both proteins can be co-purified from *E. coli* lysate with nickel-NTA resin. When 5-helix-His6 and Sac7d are co-expressed in *E. coli*, we observe no purified protein following incubation with nickel-NTA resin and elution with imidazole (Figure 5.2C, lane 1), indicating that Sac7d doesn't appreciably bind 5-helix-His6. However, when 5-helix-His6 is expressed concomitantly with Sac7d-Cpep, both proteins readily co-purify following incubation with nickel-NTA agarose beads and elution with imidazole (Figure 5.2C, lane 2), indicating that the two proteins form a tight association in a complex biological solution that contains thousands of alternative proteins, and a virtual sea of

nucleic acids and other biological molecules. Formation of the desired complex was further assessed by *in cellulo* ELISA, which indicates binding between Sac7d-Cpep and immobilized 5-helix, but no appreciable affinity between Sac7d and 5-helix (Figure 5.2D).

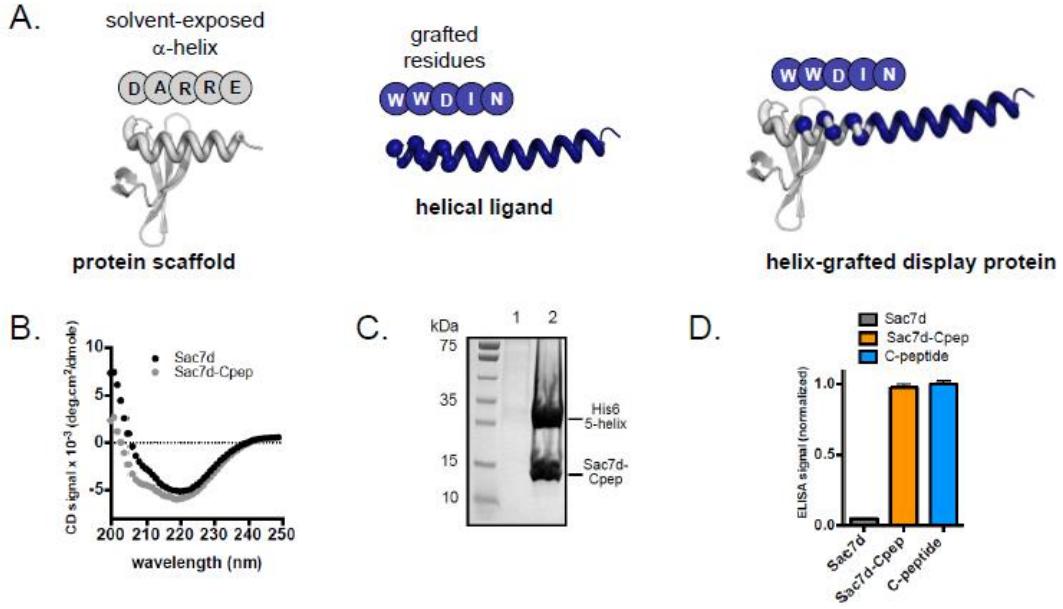


Figure 5.2 (A) Helix-grafted display of gp41 C-peptide (density blue) onto the solvent-exposed helix of a PH-like domain scaffold, Sac7d (grey). Grafted residues are shown as density blue spheres. (B) Circular dichroism data for Sac7d and Sac7d-Cpep showing retention of structural elements post-grafting. (C) Sac7d-Cpep binds 5-helix in E. coli. Lane 1: E. coli cell lysate containing Sac7d and 5-helix-His6. Sac7d does not co-purify with 5-helix-His6 on nickel-NTA agarose. Lane 2: E. coli cell lysate with Sac7d-Cpep and 5-helix-His6. Sac7d-Cpep co-purifies with 5-helix-His6 on nickel-NTA agarose. (D) *In cellulo* ELISA showing 5-helix binding to helix-grafted Sac7d-Cpep and Cpeptide in a complex cellular environment and no binding of the wildtype Sac7d scaffold to 5-helix.

5.3 Sac7d-Cpep potently inhibits HIV entry in a live virus assay.

To measure inhibition of HIV entry, or lack thereof, we previously used a live-virus assay that links HIV infection of CD4-positive T cells to GFP fluorescence.[15, 23] In this method, HIV IIIB is administered to CD4+ T cells stably integrated with a plasmid that encodes the HIV-1 long-terminal repeat (LTR) upstream of green fluorescent protein (GFP).[24] Since HIV Tat/TAR-dependent transcription ultimately leads to the expression of GFP, cells that express GFP correlates

to the number of infected cells and can thus be measured via flow cytometry. Sac7d does not inhibit HIV entry (Figure 5.3). In contrast, Sac7d-Cpep potently inhibits HIV entry (IC_{50} ~1.9-12.4 nM , Table 5.7, entry 1), and compares favorably to Enfuvirtide (IC_{50} ~0.7-3.5 nM , Table 5.7, entry 5) and gp41 C-peptide (IC_{50} ~1.0-6.7 nM , Table 5.7, entry 4). In addition to Sac7d-Cpep, we prepared and evaluated a number of other helix-grafted proteins, in which the helix consists of the Enfuvirtide sequence, or C-46, which contains C-peptide and Enfuvirtide (Figure 5.3 and Table 5.7, entries 2-3). While these proteins express as soluble proteins from *E. coli*, they are less active as HIV entry inhibitors (Sac7d-C46 IC_{50} ~23.6 nM and Sac7d-Enfuvirtide 39.7 nM). Moreover, since Enfuvirtide lacks residues (tryptophans) necessary for 5-helix recognition, this scaffold cannot be evolved for improved 5-helix recognition. Given all of this, we focused on Sac7d-Cpep as an initial starting point to generate a therapeutic cocktail from the evolution of solvent-exposed helix residues and retention—or improvement—of affinity for 5-helix.

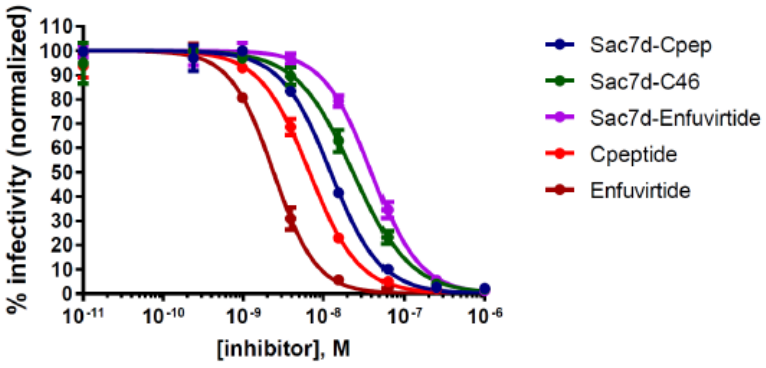


Figure 5.3 Sac7d-Cpep potently inhibits HIV entry in a live virus assay (IC_{50} ~1.9-12.4 +/- 0.2-1.3 nM). Error bars represent the standard error of the mean (SEM) for three separate experiments. If error bars are not visible, error is smaller than the data point.

5.4 First-generation helix evolution provides potent entry inhibitors with dramatically improved expression in *E. coli*.

We set out to use protein evolution to identify a cocktail of Sac7d-Cpep based proteins with affinity for 5-helix. We envisaged that if evolved forms of Sac7d-Cpep—with sequence diverse residues on the solvent-exposed helix face—retain affinity 5-helix, the molecular underpinnings of recognition could vary substantially. If true, we reasoned that it would likely be more difficult for HIV to rapidly evolve resistance to the polyclonal nature of the ensemble, compared to a single molecule like Enfuvirtide. Moreover, we reasoned that evolved proteins could bind viral gp41 in a manner that overcomes mutations leading to viral resistance.

The first two C-peptide grafted residues on Sac7d-Cpep are both tryptophans (WW, white spheres, Figure 5.4A). Multiple labs, including our own, have shown that these two tryptophan residues are critical to binding a complex between C-peptide and gp41 5-helix.[21] Thus, we avoided mutation of these critical residues. In the first-generation of helix optimization, we mutated five residues downstream of the tryptophans on the solvent-exposed face of the helix (EIYTI, light blue spheres, Figure 5.4A) to all proteinogenic amino acids, generating a protein library of approximately 3.2 x 10^6. Yeast were induced to display the Sac7d-Cpep based library, equipped with an N-terminal myc tag.[21, 25] Following display, yeast was concomitantly incubated with a FITC-labeled anti-myc antibody and biotinylated gp41 5-helix. Following a brief washing step, streptavidin-phycoerythrin conjugate (SAV-PE) was added. Thus, yeast featuring biotin near their cell surface, by virtue of a surface protein/5-helix interaction, will be bound by SA-PE (Figure 5.4B). Following additional washing, yeast were analyzed by flow cytometry, and sorted for the highest FITC fluorescence intensity (display efficiency) and PE red pigment intensity (5-helix binding efficiency). Yeast displaying tightest affinity proteins were enriched over the

course of two rounds of screening against 1 nM or 100 pM biotinylated 5-helix, respectively. In both rounds, 10 nM Cpep-ELMO was added as a competitor for 5-helix recognition (Figure 5.4C, Supplemental Table 6.4).

After two rounds of first-generation optimization we sequenced 30 clones, revealing 14 unique sequences. The amino acid preference for each randomized position is depicted in Figure 5.4D. For the first randomized residue (glutamic acid, E, Figure 5.4A), we observe a preference for the wild-type residue. Aspartic acid was also observed in selected proteins, suggesting that a carboxylic acid side chain is necessary at this position. However, we observe significant heterogeneity at the four other residues optimized by yeast display – indicating that multiple solutions to potent recognition of 5-helix exist. While the overwhelming majority of the amino acids we enriched for are hydrophobic, the steric footprint of these residues varies dramatically.

New 5-helix binders were first evaluated by *in vitro* ELISA (Figure 5.4E) and seven proteins were identified as comparing favorably with Sac7d-Cpep (greater than 90 % of ELISA signal relative to Sac7d-Cpep). These proteins were next evaluated for their ability to inhibit HIV entry. Satisfyingly, all seven proteins inhibit HIV entry virtually identically (Figure 5.4F, Table 5.7, entry 6-12). To differentiate these proteins, we measured their expression yield in *E. coli* (Figure 5.4G, Supplemental Figure 6.4B). Most dramatically, we observed an 18-fold increase in expression of Sac7d-Cpep1.1, compared to Sac7d-Cpep. First-generation residues optimized in Sac7d-Cpep1.1 are EAWLL, which vary substantially from the wild-type residues EIYTI. We wanted to characterize the improved binding interaction for the evolved C-peptide mimic. Although reports of the dissociation constants (K_D) vary for the C-peptide/5-helix complex depending on the method used to obtain the value, the K_D can be compared in each individual method to extract fold change from C-peptide mimics to the C-peptide. Using a previously reported

method, we utilized yeast display to characterize the K_D for complexes involving 5-helix and Sac7d-Cpep, Sac7d-Cpep1.1, or C-peptide. The binding interaction to 5-helix exhibited 3-fold improvement from C-peptide to Sac7d-Cpep and 7-fold improvement from C-peptide to Sac7d-Cpep1.1 (Figure 5.4H). On the basis of these collective findings, Sac7d-Cpep1.1 was used as a starting point for second-generation helix optimization.

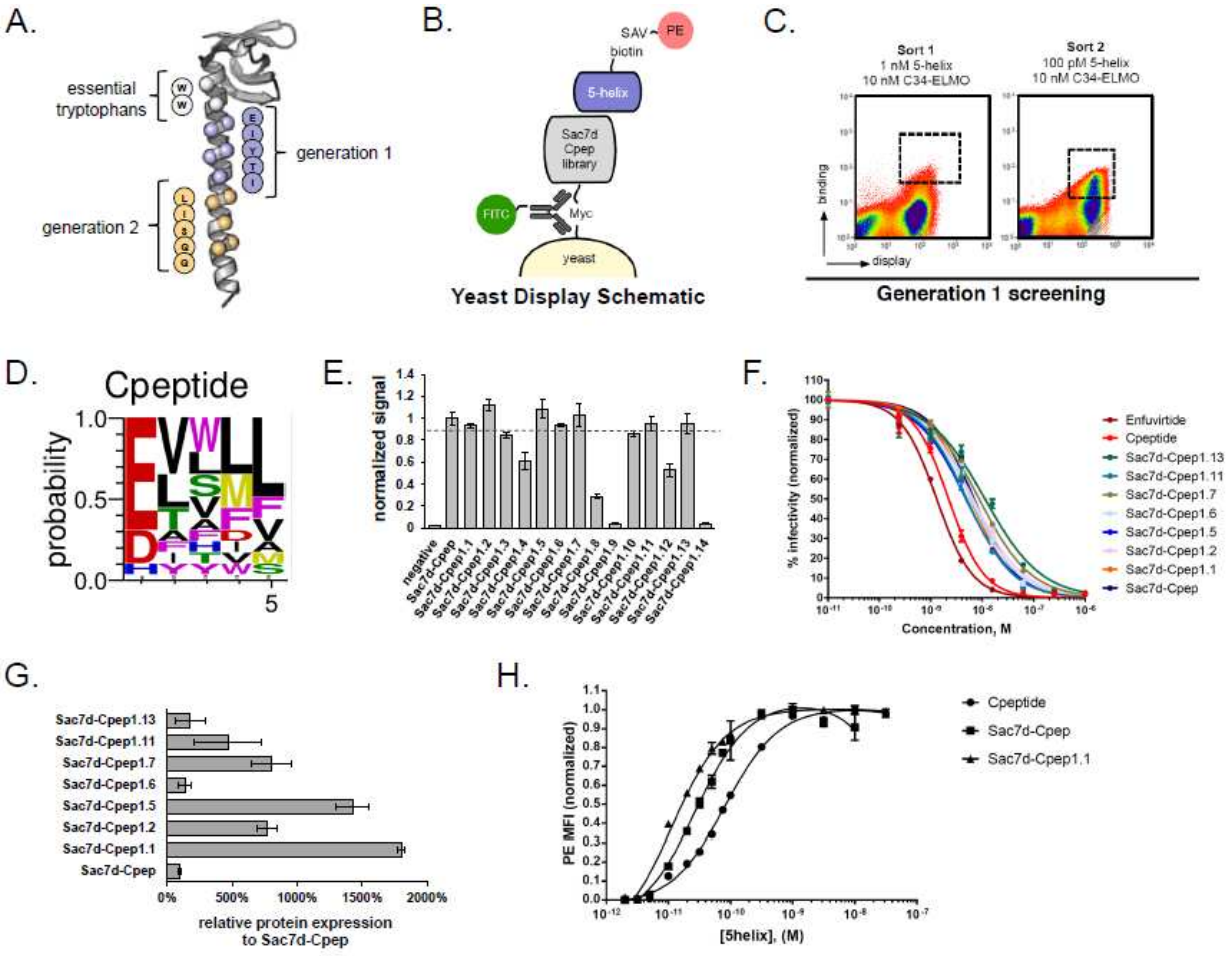


Figure 5.4 (A) Helix optimization strategy. White spheres (tryptophans) remain unaltered grafted residues due to the essential nature involved in binding gp41 5-helix. Light blue spheres were optimized in generation 1; tan spheres were optimized in generation 2. (B) Yeast display optimization schematic. (C) Yeast display data from two rounds of generation 1 screening. (D) Sequence logo for residues of 14 unique proteins optimized during generation 1 of yeast display screening. (E) *In vitro* ELISA data of generation 1 variants relative to Sac7d-Cpep signal. Dashed line represents 90% cutoff for variants progressing through further analysis. (F) Inhibition of HIV entry of Sac7d-Cpep, C-peptide, Enfuvirtude, and generation 1 optimized proteins. (G) Protein expression of generation 1 evolved proteins in *E. coli* relative to Sac7d-Cpep. (H) Binding affinity for 5-helix recognition was quantified for Cpeptide (82.6 +/- 8.6 pM), helix-grafted Sac7d-Cpep (27.4 +/- 3.7 pM) and evolved Sac7d-Cpep1.1 (11.4 +/- 2.1 pM) by displaying each protein or peptide on yeast. Varying concentrations of biotin-5-helix / streptavidin-phycoerythrin were incubated with the yeast and the normalized mean fluorescence intensity of phycoerythrin was plotted against biotin-5helix concentrations. Error bars represent the standard error of the mean (SEM) for three separate experiments. If error bars are not visible, error is smaller than the data point.

5.5 Second-generation helix evolution delivers a cocktail of potent entry inhibitors.

Using Sac7d-Cpep1.1 as a starting point, solvent-exposed residues on the C-terminal region of the grafted helix (LISQQ, tan spheres, Figure 5.4A) were randomized and enriched for 5-helix recognition by yeast display. As before, these five residues were randomized, generating a 3.2 x 10^6 protein library and highest affinity binders to 5-helix were enriched by flow cytometry. Second-generation helix optimization was conducted over the course of three rounds (4 nM; 200 pM, or 50 pM biotinylated 5-helix with 1 nM Sac7d-Cpep as competitor) (Figure 5.5A and Supplemental Table 6.4).

Following the final enrichment round, we sequenced 50 plasmids encoding evolved proteins, providing 20 unique protein sequences. We observed high levels of sequence variability at positions randomized and optimized in our yeast display screen (Figure 5.5B), suggesting that multiple solutions to potent 5-helix recognition exist within this region of the C-peptide helix. Affinity for 5-helix recognition was again assessed by *in vitro* ELISA (Figure 5.5C) and lowest affinity binders were triaged (signal less than Sac7d-Cpep). Satisfyingly, despite dramatic resurfacing of the solvent-exposed face of C-peptide, many proteins evolved in second-generation helix optimization expressed more efficiently in *E. coli*, compared to Sac7d-Cpep. The evolved proteins with low expression in *E. coli* were also eliminated (Figure 5.5D and Supplemental Figure 6.4C). Entry inhibition was measured for seven second-generation proteins (Figure 5.5E). Despite substantial differences in their sequences, all evaluated proteins potently inhibit HIV entry (IC_{50} ~4.1-29.0 nM, Table 5.7, entry 13-19).

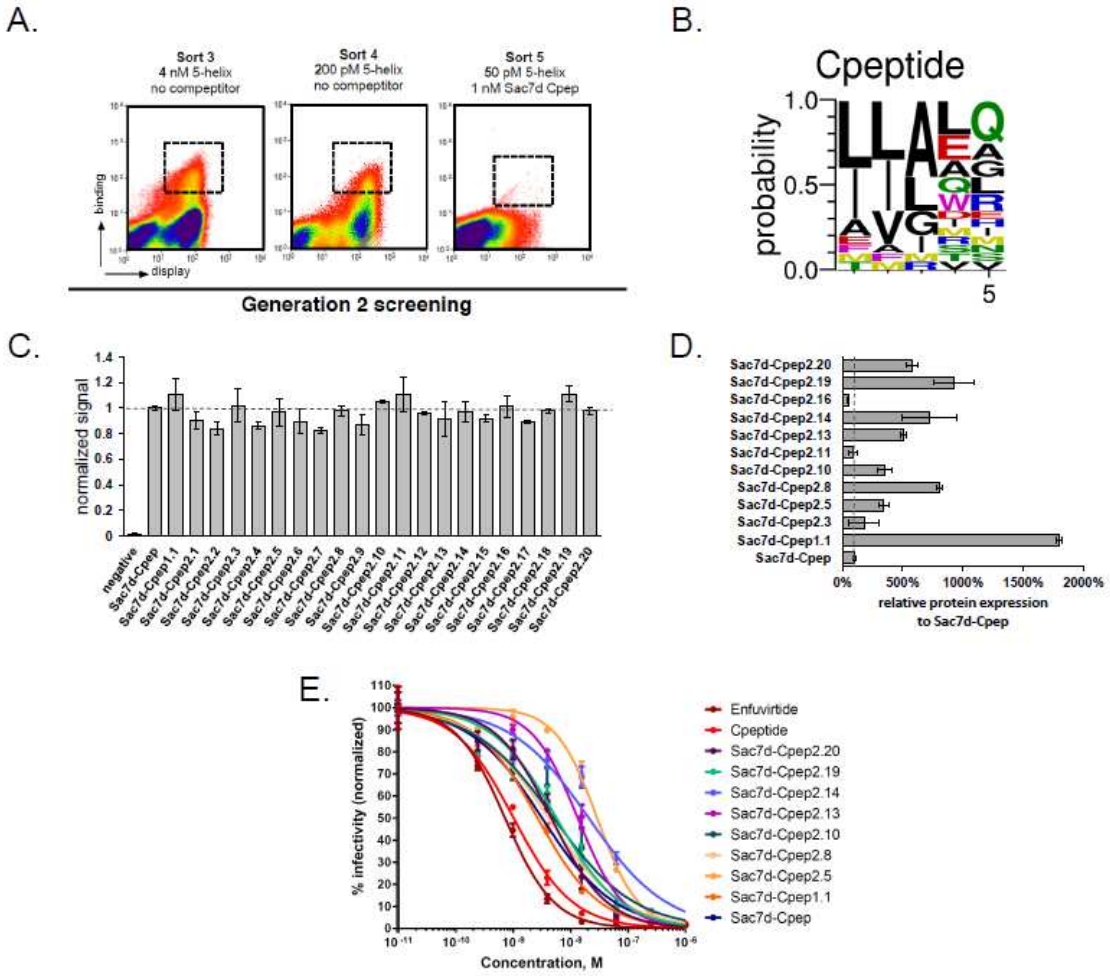


Figure 5.5 (A) Yeast display data from three rounds of generation 2 screening. (B) Sequence logo for residues of 20 unique proteins optimized during generation 2 of yeast display screening. (C) *In vitro* ELISA data of generation 2 variants relative to Sac7d-Cpep signal. Dashed line represents 100% cutoff for variants progressing through further analysis. (D) Protein expression of generation 2 evolved proteins in *E. coli* relative to Sac7d-Cpep.Dashed line represents cutoff for proteins progressing through further analysis. (E) Inhibition of HIV entry of Sac7d-Cpep, C-peptide, Enfuvirtude, and generation 1 optimized proteins. Error bars represent the standard error of the mean (SEM) for three separate experiments. If error bars are not visible, error is smaller than the data point.

5.6 Evolved proteins bind Enfuvirtide-resistant 5-helix and inhibit entry of Enfuvirtide-resistant HIV.

Given the amino acid diversity of the evolved helix, we reasoned that (at least some of) the evolved proteins recognize 5-helix (and thus gp41) in slightly different ways, or potentially with

different contributions to the binding energy at each evolved position. Given this, we rationalized that some of these evolved proteins could, in principle, inhibit mutated HIV-1 that is resistant to Enfuvirtide.

Enfuvirtide is given to patients as a last resort therapy due to *in vivo* degradation and recycling as well as rapid resistance from the virus against the peptide.[11] In patient studies, a frequent mutation to the gp41 NHR region that bestows Enfuvirtide resistance is V38A[10], however double substitutions have also been found that confer cross-resistance to Enfuvirtide and C-peptide (V38A/N42D).[26] With this in mind, we generated Enfuvirtide-resistant HIV IIIB virions with either the V38A single substitution (Figure 5.6A) or V38A/N42D double substitution (Figure 5.6B) to examine the efficacy of the evolved proteins in cocktail therapy. As in the live virus assay of the wildtype virus, HIV IIIB V38A or HIV IIIB V38A/N42D infection of CD4-positive T cells is linked to GFP fluorescence and can be quantified using flow cytometry.

We envisaged that second generation proteins would have the greatest impact on inhibition of Enfuvirtide-resistant strains based on the location of the evolved residues to the substitutions on NHR in the binding interaction (Figure 5.6A and B, generation 2). All the second-generation helix-grafted proteins exhibited potent inhibition of the Enfuvirtide-resistant double mutant strain compared to Enfuvirtide with IC_{50} values ranging from 4.3 nM to 83.7 nM which is better than Cpeptide for some mutants and comparable to the efficacy of Cpeptide for other versions (IC_{50} ~72.9 nM) (Figure 5.6C, and Table 5.7). A similar trend was seen for the single mutant strain (Figure 5.6D, and Table 5.7). Interestingly, Sac7d-Cpep and Sac7d-Cpep1.1 also retained inhibitory potency against both strains with IC_{50} values remaining relatively unchanged from wildtype HIV IIIB to HIV IIIB V38A/N42D despite the 73-fold change of C-peptide potency and 573-fold change of Enfuvirtide potency against the double substituted virus (Figure 5.6C and D

and Table 5.7, entries 1 and 6). The binding contribution from the helix-grafted Sac7d-Cpep that allows it to retain affinity despite the identical binding interface to C-peptide could be attributed to the stability of the tertiary complex for helical formation and the contribution from the native residues surrounding the grafted mutations on the solvent-exposed helix.

The diversity of amino acid sequence along the evolved helix has shown to recognize and inhibit membrane fusion of three HIV-1 strains to varying degrees. A mixture of these C-peptide mimics could be reasoned to inhibit wildtype HIV-1 and prevent rapid resistance with their polyclonal nature. To make these protein therapeutics rival mAb therapies the *in vivo* stability needs to be addressed.

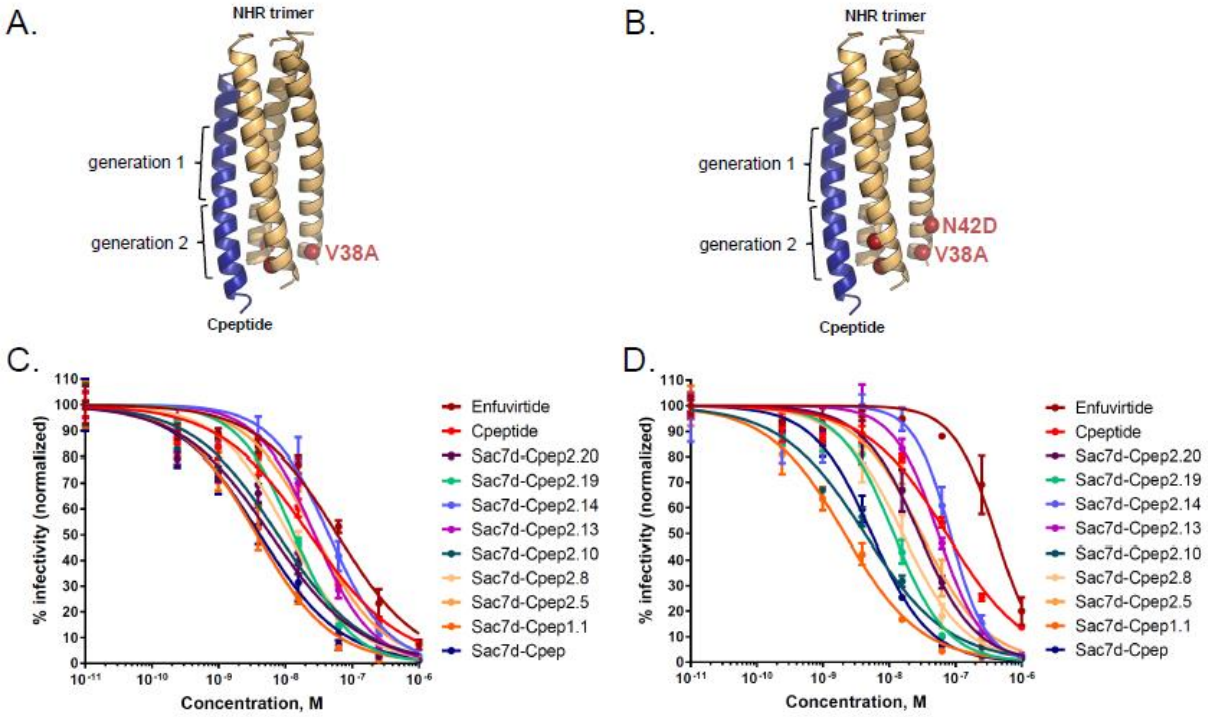


Figure 5.6 (A) Position of V38A single mutation within the N-heptad repeat (tan) of gp41 that confers Enfuvirtide resistance and (B) position of V38A and N42D double mutations within the N-heptad repeat (brown) of gp41 that confers Enfuvirtide resistance and Cpeptide sensitivity. Mutations are highlighted in red. C-peptide is highlighted in density blue; regions evolved generationally are shown. (C) Inhibition of V38A and (D) V38A/N42D HIV entry. Error bars represent the standard error of the mean (SEM) for three separate experiments. If error bars are not visible, error is smaller than the data point.

5.7 Fusions designed for serum stability potently inhibit HIV entry.

Small proteins like Sac7d-Cpep, and variations thereof, are typically cleared rapidly from the body[27, 28], compared to antibodies (e.g., IgG), which have much longer lifetimes *in vivo* (ca. half-life is approximately 22 days). This incredible stability *in vivo* is largely due to an interaction between the neonatal receptor (FcRn) on the surface of epithelial cells, and an epitope largely within the CH3 domain of a single chain of the fragment crystallizable region (Fc) (Figure 5.7A).[29, 30] This single chain is referred to as monomeric Fc (mFc, Figure 5.7A); a single CH3 domain is referred to as monomeric CH3 (mCH3, Figure 5.7A). Binding between mFc and FcRn is pH-dependent. Formation of the Fc/FcRn complex leads to a complicated process that continuously shuttles the antibody from the circulatory system to the cell interior. Once inside an endosome, which has a lower pH compared to the cytosol, the complex dissociates and the antibody is shipped back to the circulatory system.[29-31] This biological shell game allows antibodies to evade cellular degradation. Researchers have shown that fusion of a protein to mFc, or mCH3 can dramatically improve *in vivo* stability, and mFc or mCH3 fusions can often be expressed in *E. coli*.

We prepared two fusions, mFc-Sac7d-Cpep and mCH3-Sac7d-Cpep (Figure 5.7B). Given that Sac7d binds IgG, all protein fusions had an L33T mutation to turn 'off' affinity for IgG.[22] *In vitro* ELISA experiments indicate no appreciable binding between Sac7d L33T and either mFc or mCH3 (Supplemental Figure 6.4D). mCH3-Sac7d-Cpep (24 kDa) compares favorably to Sac7d-Cpep, but the increased size results in a 1.5-fold decrease in entry inhibition (IC_{50} ~18.0 nM, Figure 5.7C and Table 5.7, entry 22). Consistent with previously observed relationships between protein size and entry inhibition potency, mFc-Sac7d-Cpep (37 kDa) retains modest inhibition of HIV entry (IC_{50} ~282.9 nM, Figure 5.7D and Table 5.7, entry 20), but is approximately 23-fold less

potent than Sac7d-Cpep. Additionally, mCH3-Sac7d-Cpep and mFc-Sac7d-Cpep proteins containing a (GGS)3 linker between the mFc or mCH3 and Sac7d-Cpep performed similarly to proteins without a linker, despite the increase in size (Figure 5.7C and D and Table 5.7, entries 21 and 23).

Fusion to a polypeptide that binds serum albumin has also been reported as a strategy to improve *in vivo* stability of biologics – presumably because binding to serum albumin effectively shields proteins from serum proteases and shuttles proteins through the FcRn pathway (Figure 5.7A).[29, 31-33] We therefore prepared a fusion consisting of a 20-amino acid serum albumin binding peptide (SAbp) and Sac7d-Cpep (SAbp-Sac7d-Cpep, Figure 5.7B). Again, consistent with previously observed relationships between protein size and entry inhibition potency, SAbp-Sac7d-Cpep (14 kDa, IC_{50} ~3.9 nM) compares identically to Sac7d-Cpep (12 kDa, IC_{50} ~1.9-12.4 nM) (Figure 5.7E and Table 5.7, entry 24). Additionally, since SAbp was a minimalist fusion, we tried fusing it to the C-terminus of Sac7d-Cpep. Concerned this would unfold the helix, we inserted a helix-forming linker between Sac7d-Cpep and SAbp.[33] This fusion protein (Sac7d-Cpep-hfl-SAbp) showed similar inhibition (IC_{50} ~9.7 nM, Figure 5.7E and Table 5.7, entry 25).

Collectively, we report a new helix-grafted display biologic, Sac7d-Cpep, that potently inhibits the entry of HIV-1 in CD4-positive mammalian cells (IC_{50} ~1.9-12.4 nM). Yeast display evolution of native and grafted amino acids on the solvent-exposed helix face led to the generation of a cocktail of HIV-1 entry inhibitors with virtually identical potencies, despite significant sequence variation. While hydrophobic residues are principally selected for, the size of evolved residue side chains varies dramatically. Importantly, evolved biologics potently inhibit the entry of mutated forms of HIV-1 containing Enfuvirtide-resistant changes in amino acid composition in gp41. Helix-grafted HIV-1 entry inhibitors fused to either mCH3, mFc, or SAbp, designed for

serum stability and long *in vivo* residence, retain the ability to inhibit HIV-1 entry; however, we do observe a correlation between the size of the fusion and potency – larger fusions generally have lower potency. However, all proteins reported here potently inhibit HIV-1 entry, including entry of Enfuvirtide-resistant HIV-1 (IC_{50} ~2.1-83.7 nM, Table 5.7). Given the amino acid diversity of evolved biologics we report, within the solvent-exposed helix face, we speculate that these biologics bind the C-peptide binding cleft of 5-helix somewhat differently, with varied energetic contributions at each evolved amino acid position. Thus, we also speculate that when used as a cocktail, it might be unlikely that HIV-1 is able to rapidly evolve resistance against the entire therapeutic cocktail.

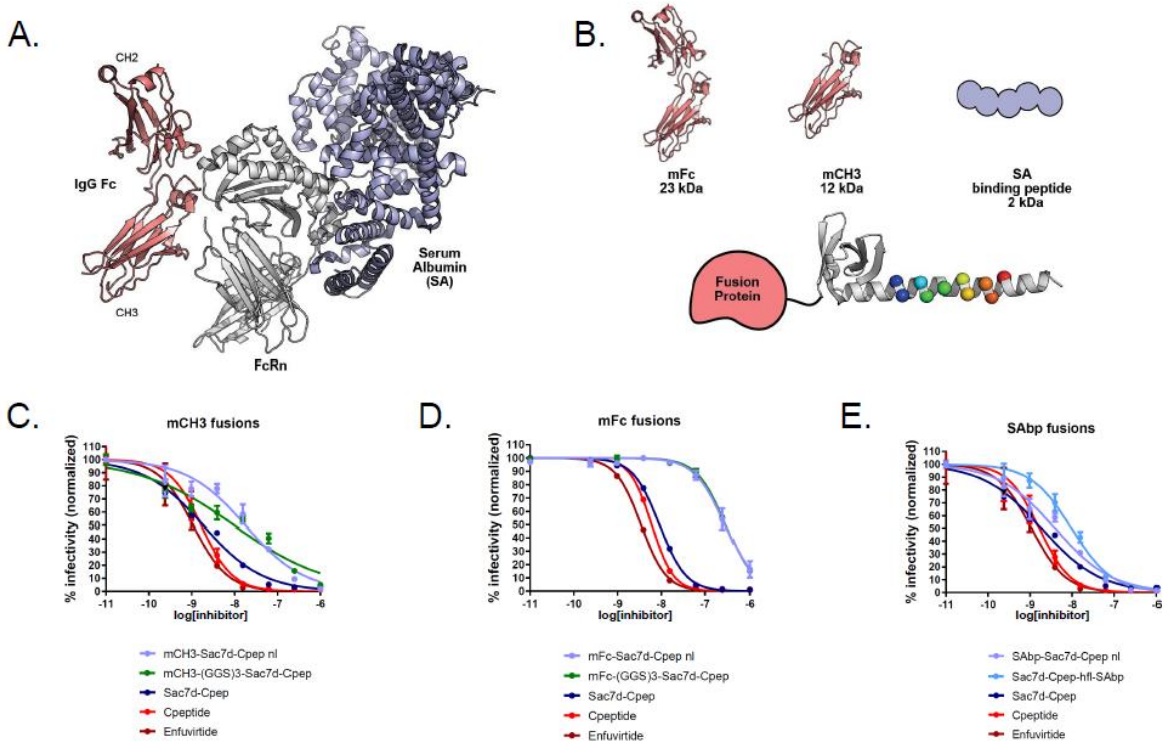


Figure 5.7 (A) Crystal structure of the FcRn receptor (grey) bound to both mFc (light pink) and Serum Albumin (light blue) (B) Fusion protein candidates (mFc ~21 kDa, mCH3 ~12kDa, SAbp ~2kDa) to elongate *in vivo* stability. (C) Inhibition data of fusion protein mCH3-Sac7d-Cpep, (D) mFc-Sac7d-Cpep, or (E) SAbp-Sac7d-Cpep with and without linkers. Error bars represent the standard error of the mean (SEM) for three separate experiments. If error bars are not visible, error is smaller than the data point.

Table 5.7 Sequence of HIV entry inhibitor biologics and associated half-maximal inhibitory concentration (IC_{50}) values for all live virus assays. Range represents cumulative data from all experiments of a given protein. Error represents the standard error of the mean (SEM) and the range of associated experiments.

				generation 1					generation 2					HIV IIIB IC_{50} (nM)	V38A IC_{50} (nM)	V38A/N42D IC_{50} (nM)
1	Sac7d-Cpep	W	W	E	I	Y	T	I	L	I	S	Q	Q	1.9-12.4 +/- 0.2-1.3	4.1 +/- 0.6	5.4 +/- 0.4
2	Sac7d-C46	W	W	E	I	Y	T	I	L	I	S	Q	Q	23.6 +/- 1.4	-	-
3	Sac7d-Enfuvirtide					Y	T	I	L	I	S	Q	Q	39.7 +/- 1.2	-	-
4	Cpeptide	W	W	E	I	Y	T	I	L	I	S	Q	Q	1.0-6.7 +/- 0.1-0.2	23.2 +/- 4.1	72.9 +/- 12.4
5	Enfuvirtide					Y	T	I	L	I	S	Q	Q	0.7-3.5 +/- 0.1-0.2	60.0 +/- 11.6	401.3 +/- 47.4
6	Sac7d-Cpep1.1	W	W	E	A	W	L	L	L	I	S	Q	Q	2.6-6.2 +/- 0.3-0.6	3.7 +/- 0.4	2.1 +/- 0.2
7	Sac7d-Cpep1.2	W	W	E	Y	L	I	L	L	I	S	Q	Q	7.7 +/- 1.4	-	-
8	Sac7d-Cpep1.5	W	W	E	I	A	L	V	L	I	S	Q	Q	5.4 +/- 0.9	-	-
9	Sac7d-Cpep1.6	W	W	E	T	V	M	L	L	I	S	Q	Q	7.0 +/- 1.2	-	-
10	Sac7d-Cpep1.7	W	W	E	L	S	W	L	L	I	S	Q	Q	9.4 +/- 1.4	-	-
11	Sac7d-Cpep1.11	W	W	D	V	L	L	M	L	I	S	Q	Q	5.8 +/- 1.0	-	-
12	Sac7d-Cpep1.13	W	W	E	L	Y	F	S	L	I	S	Q	Q	15.6 +/- 4.0	-	-
13	Sac7d-Cpep2.5	W	W	E	A	W	L	L	I	V	L	T	L	29.0 +/- 3.5	27.7 +/- 5.7	30.5 +/- 8.7
14	Sac7d-Cpep2.8	W	W	E	A	W	L	L	L	F	L	A	Q	4.1 +/- 0.8	11.2 +/- 1.7	15.9 +/- 2.3
15	Sac7d-Cpep2.10	W	W	E	A	W	L	L	L	L	M	Q	Q	5.1 +/- 1.5	7.7 +/- 1.8	4.3 +/- 0.6
16	Sac7d-Cpep2.13	W	W	E	A	W	L	L	L	I	G	M	S	13.3 +/- 3.4	27.4 +/- 4.2	55.0 +/- 7.0
17	Sac7d-Cpep2.14	W	W	E	A	W	L	L	I	L	I	E	I	20.7 +/- 7.9	44.8 +/- 8.5	83.7 +/- 16.7
18	Sac7d-Cpep2.19	W	W	E	A	W	L	L	L	L	A	R	H	5.7 +/- 1.1	13.5 +/- 1.6	11.7 +/- 1.8
19	Sac7d-Cpep2.20	W	W	E	A	W	L	L	F	L	A	S	Q	4.6 +/- 0.6	6.0 +/- 1.5	28.2 +/- 4.1
20	mFc-Sac7d-Cpep nl	W	W	E	I	Y	T	I	L	I	S	Q	Q	282.9 +/- 17.8	-	-
21	mFc-GGS3-Sac7d-Cpep	W	W	E	I	Y	T	I	L	I	S	Q	Q	291.5 +/- 9.9	-	-
22	mCH3-Sac7d-Cpep nl	W	W	E	I	Y	T	I	L	I	S	Q	Q	18.0 +/- 4.8	-	-
23	mCH3-GGS3-Sac7d-Cpep	W	W	E	I	Y	T	I	L	I	S	Q	Q	9.8 +/- 3.8	-	-
24	SAbp-Sac7d-Cpep nl	W	W	E	I	Y	T	I	L	I	S	Q	Q	3.9 +/- 0.8	-	-
25	Sac7d-Cpep-hfl-SAbp	W	W	E	I	Y	T	I	L	I	S	Q	Q	9.7 +/- 1.1	-	-

5.8 Conclusion

HIV/AIDS afflicts 36.7 million people worldwide, and currently there is no vaccine or cure. Current HIV drugs target many facets of the virus lifecycle; however, entry inhibitors are limited and come with their own challenges. Moreover, most current HIV drugs require daily administration. Here, we set out to develop gp41 C-peptide helix-grafted proteins that potently bind gp41, inhibit HIV entry, and are designed to have *in vivo* stabilities that compare favorably to monoclonal antibodies. Expanding on our prior work using helix-grafted display, here we report Sac7d-Cpep, which potently inhibits HIV entry in a live virus assay (IC_{50} ~1.9-12.4 nM). Using yeast display, we resurfaced the solvent-exposed face of the grafted helix. This resulted in numerous new proteins with sequence-diverse grafted helices, which potently inhibit HIV entry. Given the sequence diversity of these helices, we speculated that these evolved proteins likely bind

gp41 slightly differently, and therefore might be effective against HIV strains that are resistant to Enfuvirtide, an FDA approved HIV entry inhibitor that binds prefusogenic gp41. Satisfyingly, we found that many of our evolved proteins inhibit entry of an HIV strain resistant to Enfuvirtide. Finally, we demonstrated that protein fusions consisting of the helix-grafted display entry inhibitor and polypeptides shown to increase in vivo stability express readily in *E. coli*, and in some cases, exhibit properties (recognition of the neonatal receptor) necessary for extended in vivo stability. Collectively, we report a virtual cocktail of HIV entry inhibitors that potently block HIV entry—including in a strain resistant to Enfuvirtide—and demonstrate function necessary for long in vivo lifetime. Future work associated with this research would analyze the in vivo stability using animal models to produce a therapeutic with minimal immunogenic effect and maximal inhibitory effects.[23]

5.9 Methods

Protein Expression and Purification

Genes were cloned into pET using BamHI and KpnI, downstream of a His6 tag and transformed into BL21s (DE3). Cells were grown in 0.5 L LB cultures containing 100 μg/mL carbenicillin at 37 °C to OD600 =0.5 - 0.8 and induced with 1 μM IPTG at 25 °C overnight. Cells were then collected by centrifugation, resuspended in PBS buffer (20 mM Na2HPO4 pH 7.4, 100 mM NaCl) with protease tablets and stored at -20 °C. Frozen pellets were thawed and sonicated with 1 sec pulses for 2 min. The lysate was cleared by centrifugation (11,899 xg, 10 min.) and the supernatant was mixed with 1 mL of Ni-NTA agarose resin for 30 min at 4 °C. The resin was collected by centrifugation (4,303 xg, 10 min.). The resin was sequentially washed with 30 mL of PBS buffer containing 20 mM imidazole, 50 mL PBS buffer containing 50 mM imidazole, and 15 mL PBS

buffer containing 75 mM imidazole. The protein was then eluted with 4 mL PBS buffer containing 400 mM imidazole. The proteins were dialyzed against PBS buffer and analyzed for purity by SDS-PAGE. Purified protein concentrations were quantified using Beer's Law at an absorbance of 280 nm, following standard practice. Protein expression was calculated as a measure of eluted protein yield in milligrams per liter of induced E. coli culture. Protein fusions with either mFc or mCH3 were grown up in 2XYTand induced at 20 °C for 15-20 hours for maximal yield.

Circular Dichroism

Proteins were purified as described above. Separately, each protein was diluted to 5-12 μM in PBS buffer. Wavelength data are the average of three scans from 250 nm to 200 nm in 1 nm steps at 25 °C. Thermal denaturation experiments at 222 nm were run from 0 to 90 °C in two-degree steps at a two-degree/min rate of increase with one min equilibration and data averaging at each temperature. Tm values were obtained from minima of the first derivative of θ versus 1/T plots.

Cell lysate ELISA

Sac7d, Sac7d-Cpep, and C-peptide were cloned into MCS1 of pETDuet-1 with FLAG tags using NcoI and NotI. The 5-Helix with a C-terminal His6 tag was cloned into MCS2 of pETDuet-1 using NdeI and KpnI. For protein expression, constructs were transformed into BL21 cells (DE3). Cells containing the co-expressed pair were inoculated and induced in 10 mL LB cultures overnight. Cells were spun down and resuspended in 10 mL PBS buffer, lysed by sonication, and spun down to remove cell debris. Cleared lysates were incubated on clear Ni-NTA coated plates for 1 hr at room temperature and each well was washed 3x with 200 μL wash buffer (PBS Buffer, 0.1% Tween® 20, and 0.01 mg/mL BSA). An HRP-conjugated mouse anti-DDDDK antibody in LiCor Blocking Buffer was incubated for 1 hr at room temperature, followed by 3 washes. Color was

developed using TMB-One substrate and absorbance was measured at 655 nm on a SynergyMx Microplate Reader.

Plasmid Construction of HIV-1 Infectivity Assays

HIV-1 IIIB C200 proviral expression construct has been described previously. NheI/BamHI fragments of the env gene encoding substitutions (gp41 V38A or gp41 V38A N42D) were synthesized (INTEGRATED DNA TECHNOLOGIES), cloned into pJET1.2/blunt cloning vector (Thermo Scientific) by amplifying with the following primer set (5'-GCT AGC AAA TTA AGA GAA CAA TTT GG-3' and 5'-GGA TCC GTT CAC TAA TCG AAT GG-3') and sequenced as described previously. Then, DNA fragments were inserted into the proviral DNA expression plasmid at NheI/BamHI site.

Infectivity Assay

The procedure of HIV-1 infectivity assays was previously reported. Briefly, viruses were produced by transfection of 3.0 μg of HIV-1 IIIB C200 proviral expression construct into 293T cells ($3.0x10^6$) using TransIT®-LT1 reagent. 48h later, virus-containing supernatants were filtered by 0.45 μm filters (Millipore) and used to infect into $2.5x10^4$ CEM-GFP cells with varying concentration of inhibitors including T-20 and C34. Infectivity (GFP+ cells) was measured by flow cytometry at 2 days post-infection.

Resolubilization of 5-helix Inclusion Bodies

5-helix-His6 was cloned into a modified pETDuet-1 vector using NdeI and KpnI and transformed into BL21 cells. Cells were induced to express 5-helix-His6 and lysed as described above. The lysate was cleared by centrifugation (11,899 x g, 20 min) and the supernatant was discarded. The pellet was washed twice with PBS buffer containing 0.5 % Triton® X-100 and once with PBS buffer, resuspended in urea buffer (PBS buffer with 8 M urea and 10 mM imidazole) to resolubilize

the inclusion bodies and cleared by centrifugation (11,899 xg, 30 min.) The supernatant was mixed with 1 mL of Ni-NTA agarose resin for 1 hr at 4 °C. After centrifugation (4,303 x g, 4 min), the resin was washed with 50 mL of urea buffer and eluted with 40 mL of urea elution buffer (PBS buffer with 6 M urea and 100 mM imidazole) into 460 mL PBS buffer by gravity elution while stirring to refold the protein. The 500 mL elution was run through a column containing 1 mL of Ni-NTA agarose resin and eluted with 5 mL PBS buffer containing 400 mM imidazole. The protein was dialyzed against PBS buffer and analyzed for purity by SDS-PAGE. Refolding analysis was conducted by CD. Purified proteins were quantified using Beer's Law at an absorbance of 280 nm. Protein expression was calculated as a measure of eluted protein yield in milligrams per liter of induced E. coli culture.

Biotinylation

5-helix-His6 was cloned into a pET vector containing an upstream Avitag™ (GLNDIFEAQKIEWHE)-GGSGGSGGT linker using KpnI and PacI. The protein was resolubilized from inclusion bodies as described above in PBS. His6-BirA was cloned into pET using NcoI and KpnI and purified as described above in PBS. 300 µL of Avitagged™ 5-Helix-His6 protein at 38 µM was incubated with 6 µL of His6-BirA at 1 mg/mL using Avidity® BirA biotin-protein ligase standard reaction kit at 30°C for 40 min. Biotinylation was confirmed by Agilent 6220 TOF LC-MS.

***In vitro* ELISA**

Sac7d and helix-grafted Sac7d-Cpep were cloned into a pET vector downstream of an N-terminal His6-FLAG tag (DYKDDDDK) using BamHI and Kpn1 or BamHI and PacI. Proteins were purified as described above in PBS buffer. The biotinylated Avi-tagged 5-helix-His6 (biotin-5-helix) was prepared as described above in PBS buffer and diluted to 10 µg/mL. Pierce®

Streptavidin coated clear 96-well plates with a binding capacity of 5 pmol were pre-blocked with 200 µL of wash buffer (PBS, 0.5 mg/mL BSA, 0.1% tween® 20) for 1 hr. Biotin-5-helix was immobilized on the streptavidin-coated plates by incubating 100 µL of diluted protein for 1 hr at room temperature, followed by 3 washes. 100 µL of Sac7d or Sac7d-Cpep was incubated in various concentrations for 1 hr and washed 3 times at 4°C (All concentrations of incubated helix-grafted protein were above ligand-depleting conditions for binding the 5 pmol immobilized biotin-5-helix in each well). An HRP-conjugated mouse anti-DDDDK antibody in LICOR Blocking Buffer (1:10,000) was incubated for 1 hr at room temperature, followed by 4 washes at 4°C. Color was developed using TMB-One substrate at room temperature and absorbance was measured at 655 nm on a SynergyMx Microplate Reader after 20 min.

Preparation of Protein Library Generation 1

EBY100 yeast (trp-, leu-, with the Aga1p gene stably integrated) and the pCTCON2 plasmid were generously provided by the Wittrup lab (MIT). The gene encoding Sac7d-Cpep was amplified by PCR and cloned into pCTCON2 in-frame with Aga2, an N-terminal HA-tag, and an N-terminal myc tag, using the restriction enzymes NheI and XhoI. When analyzed for display only, Sac7d-Cpep displayed efficiently on EBY100 cells (~60-85 %, data not shown). Next, the Sac7d-Cpep generation 1 library was created by amplifying the N-terminal Sac7d-Cpep gene with an N-terminal homologous recombination forward primer annealing to 40 base pairs upstream of the gene on pCTCON2 and a reverse homologous recombination primer containing 5 sites in the C-terminal α-helical region (Glu65, Ile66, Tyr69, Thr70, and Ile73) substituted with NNK codons and annealing to 40 base pairs downstream of the final NNK mutation. The resulting amplicon, containing the randomized sequences, was then cloned into a nonsense-Cpep pCTcon2 vector containing the remainder of the C-terminus of Sac7d-Cpep post-mutations by adding ~1 µg of

nonsense-Cpep pCTCON2 vector cut with NheI and NcoI mixed with ~3 µg of the amplified library, ethanol precipitated, and transformed via homologous recombination into electrocompetent EBY100. Eight of these electroporations were performed to reach appropriate transformation efficiency. Rescued yeast were resuspended in 1 mL fresh SD-CAA (5.4 g/L Na_2HPO_4, 8.6 g/L $NaH_2PO_4 \cdot H_2O$, 20 g/L dextrose, 6.7 g/L yeast nitrogen base lacking amino acids, 5 g/L casamino acids, 200 kU/L penicillin, 0.1 g/L streptomycin). A small portion was plated by serial dilution onto SD-CAA plates, and incubated at 30 °C for 3 days in order to determine the transformation efficiency. The remainder were cultured in 50 mL SD-CAA for yeast display screening. Efficiency of homologous recombination transformation was determined to be 1.072 x 10^7 – 2.19 x 10^7 using the equation transformation efficiency = (# colonies x Vol quenched)/(dilution x Vol plated).

Preparation of Protein Library Generation 2

The gene encoding the best performing mutant from generation 1, Sac7d-Cpep1.1 pCTCON2 was electroporated back into pCTCON2 in-frame with Aga2, an N-terminal HA-tag, and an N-terminal myc tag, using NheI and XhoI. When analyzed for display only, Sac7d-Cpep1.1 displayed efficiently on EBY100 cells (~60-85 %, data not shown). Next, the Sac7d-Cpep generation 2 library was created by amplifying the N-terminal Sac7d-Cpep1.1 gene with an N-terminal homologous recombination forward primer annealing to 40 base pairs upstream of the gene on pCTCON2 and a reverse homologous recombination primer containing 5 sites in the C-terminal α-helical region (Leu76, Ile77, Ser80, Gln83, and Gln84) substituted with NNK codons and annealing to 40 base pairs downstream of the final NNK mutation. The resulting amplicon, containing the randomized sequences, was then cloned into a nonsense-Cpep2 pCTCON2 vector containing the remainder of the C-terminus of Sac7d-Cpep1.1 post-generation 2 mutations using

homologous recombination in EBY100 yeast by using approximately 1 µg of nonsense-Cpep2 pCTCON2 vector cut with NheI and PacI mixed with ~3 µg of the amplified generation 2 library, ethanol precipitated, and transformed via ten electroporations. Rescued yeast plated as described previously to determine the transformation efficiency and the remainder were cultured in 50 mL SD-CAA for yeast display screening. Efficiency of homologous recombination transformation was determined to be $2.084 \times 10^7 - 4.9 \times 10^7$.

Yeast Display Screening

After 2-3 days of growth in SD-CAA, the library was sub-cultured in SD-CAA at an initial density of 0.5×10^7 cells/mL and grown to a density of 2.0×10^7 cells/mL. Yeast were subsequently sub-cultured in SG-CAA (Galactose containing induction media) to a concentration of 1.0×10^7 cells/mL and grown for 1-2 days shaking at 250 rpm at 25 °C. For each round of screening, approximately 10^8 cells were pelleted and washed with 1 mL of 4 °C PBS-BSA (PBS buffer with 1 g/L BSA filter sterilized). Yeast were subsequently incubated with biotin-5-helix at the concentrations given in Supplemental Table 6.4 and a fluorescein isothiocyanate (FITC)-conjugated anti-myc antibody (1:250) at room temperature. After incubation, the yeast cells were incubated on ice for 5 min, pelleted at 12,000 xg for 30 sec 4 °C and washed with 1 mL ice-cold PBS-BSA. The yeast were pelleted again and incubated with a 1:100 dilution of streptavidin, R-Phycoerythrin (SA-PE) in PBS-BSA on ice for 1 hr. After incubation, a final wash with ice-cold PBS-BSA was performed, and yeast positive for both FITC (display) and R-Phycoerythrin (PE, 5-helix-binding) were sorted into 7 mL of SD-CAA media using a MoFlo Flow Cytometer (Beckman-Coulter). Sorted yeast were transferred to 50 mL of pre-warmed SD-CAA and incubated at 30 °C for 3 days shaking at 250 rpm. Additionally, plasmid DNA from the sorted

library population was recovered using a Zymoprep yeast plasmid miniprep II kit. This DNA was used to transform Invitrogen, Top10 *E. coli* and analyzed for sequence diversity.

Viral Infectivity Assays for IC_{50} Determination

The yield of infected cells produced was quantified by flow cytometry and normalized relative to maximum dosages. The concentrations that caused 50% inhibition (IC_{50}) were calculated using the nonlinear fit variable slope model where n=3 (Prism, GraphPad Software). Error in IC_{50} values were determined for each concentration using standard deviation and each concentration was evaluated for significance by T-test.

Statistical Analysis

Statistical tests and the associated standard deviation error bars are identified in the figure captions. Statistical analysis was performed using Prism (GraphPad Software).

5.10 Sequences of Proteins Used in this Work

His-Sac7d

MGSSHHHHHHSQDPVKVKFKYKGEEKEVDTSKIKKVWRVGKMVSFLYDDNGKTGRG
AVSEKDAPKELLDMLARAEREKKLN*

His-Sac7d Cpep

MGSSHHHHHHSQDPVKVKFLLNGEEKEVDTSKIRDVSRQGKNVKFLYNDNGKYGAGN
VDEKDAPKELLWMLWDAEINKYTSLIHSLIEESQNQQEKNEQELL*

His-FLAG Sac7d Cpep

MGSSHHHHHHSQDPDYKDDDDKVKVKFLLNGEEKEVDTSKIRDVSRQGKNVKFLYND
NGKYGAGNVDEKDAPKELLWMLWDAEINKYTSLIHSLIEESQNQQEKNEQELL*

Cmyc Sac7d Cpep

LASEQKLISEEDLGSVKVKFLLNGEEKEVDTSKIRDVSRQGKNVKFLYNDNGKYGAGN

VDEKDAPKELLWMLWDAEINKYTSLIHSLIEESQNQQEKNEQELL

FLAG Tag

DYKDDDDK

Cmyc

LASEQKLISEEDLGS

C-peptide

WMEWDREINNYTSLIHSLIEESQNQQEKNEQELL

SAbp

QRHPEDICLPRWGCLWGDDD

mCH3

GQCREPQVYTSPPSRDELTKNQVSLRCHVKGFYPSDIAVEWESNGQPENNYKTTKPVLD

SDGSFRLASYLTVDKSRWQQGNVFSCSVMHECLHNHYTQKSLSLSPGK

mFc

APELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKT

KPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQ

VYTKPPSRDELTKNQVSLSCLVKGFYPSDIAVEWESNGQPENNYKTTVPVLDSDGSFRL

ASYLTVDKSRWQQGNVFSCSVMHEALHNHYTQKSLSLSPGK*

5-Helix-His6x

MQLLSGIVQQQNNLLRAIEAQQHLLQLTVWGIKQLQARILAGGSGGHTTWMEWDREIN

NYTSLIHSLIEESQNQQEKNEQELLEGSSGGQLLSGIVQQQNNLLRAIEAQQHLLQLTVW

GIKQLQARILAGGSGGHTTWMEWDREINNYTSLIHSLIEESQNQQEKNEQELLEGSSGGQ

LLSGIVQQQNNLLRAIEAQQHLLQLTVWGIKQLQARILAGGHHHHHH

Generation 1 library sequences

Sac7d-Cpep

VKVKFLLNGEEKEVDTSKIRDVSRQGKNVKFLYNDNGKYGAGNVDEKDAPKELLWML

WDAEINKYTSLIHSLIEESQNQQEKNEQELL

Sac7d-Cpep1.1

VKVKFLLNGEEKEVDTSKIRDVSRQGKNVKFLYNDNGKYGAGNVDEKDAPKELLWML

WDAEANKWLSLLHSLIEESQNQQEKNEQELL

Sac7d-Cpep1.2

VKVKFLLNGEEKEVDTSKIRDVSRQGKNVKFLYNDNGKYGAGNVDEKDAPKELLWML

WDAEYNKLISLLHSLIEESQNQQEKNEQELL

Sac7d-Cpep1.3

VKVKFLLNGEEKEVDTSKIRDVSRQGKNVKFLYNDNGKYGAGNVDEKDAPKELLWML

WDAEVNKWLSLFHSLIEESQNQQEKNEQELL

Sac7d-Cpep1.4

VKVKFLLNGEEKEVDTSKIRDVSRQGKNVKFLYNDNGKYGAGNVDEKDAPKELLWML

WDAHVNKSMSLLHSLIEESQNQQEKNEQELL

Sac7d-Cpep1.5

VKVKFLLNGEEKEVDTSKIRDVSRQGKNVKFLYNDNGKYGAGNVDEKDAPKELLWML

WDAEINKALSLVHSLIEESQNQQEKNEQELL

Sac7d-Cpep1.6

VKVKFLLNGEEKEVDTSKIRDVSRQGKNVKFLYNDNGKYGAGNVDEKDAPKELLWML

WDAETNKVMSLLHSLIEESQNQQEKNEQELL

Sac7d-Cpep1.7

VKVKFLLNGEEKEVDTSKIRDVSRQGKNVKFLYNDNGKYGAGNVDEKDAPKELLWML

WDAELNKSWSLLHSLIEESQNQQEKNEQELL

Sac7d-Cpep1.8

VKVKFLLNGEEKEVDTSKIRDVSRQGKNVKFLYNDNGKYGAGNVDEKDAPKELLWML

WDAEVNKHMSLVHSLIEESQNQQEKNEQELL

Sac7d-Cpep1.9

VKVKFLLNGEEKEVDTSKIRDVSRQGKNVKFLYNDNGKYGAGNVDEKDAPKELLWML

WDAETNKVLSLLHSLIEESQNQQEKNEQELL

Sac7d-Cpep1.10

VKVKFLLNGEEKEVDTSKIRDVSRQGKNVKFLYNDNGKYGAGNVDEKDAPKELLWML

WDAEFNKFDSLLHSLIEESQNQQEKNEQELL

Sac7d-Cpep1.11

VKVKFLLNGEEKEVDTSKIRDVSRQGKNVKFLYNDNGKYGAGNVDEKDAPKELLWML

WDADVNKLLSLMHSLIEESQNQQEKNEQELL

Sac7d-Cpep1.12

VKVKFLLNGEEKEVDTSKIRDVSRQGKNVKFLYNDNGKYGAGNVDEKDAPKELLWML

WDADVNKTVSLAHSLIEESQNQQEKNEQELL

Sac7d-Cpep1.13

VKVKFLLNGEEKEVDTSKIRDVSRQGKNVKFLYNDNGKYGAGNVDEKDAPKELLWML

WDAELNKYFSLSHSLIEESQNQQEKNEQELL

Sac7d-Cpep1.14

VKVKFLLNGEEKEVDTSKIRDVSRQGKNVKFLYNDNGKYGAGNVDEKDAPKELLWML

WDADLNKWFSLFHSLIEESQNQQEKNEQELL

Generation 2 library sequences

Sac7d-Cpep2.1

VKVKFLLNGEEKEVDTSKIRDVSRQGKNVKFLYNDNGKYGAGNVDEKDAPKELLWML

WDAEANKWLSLLHSLVEEAQNWEEKNEQELL

Sac7d-Cpep2.2

VKVKFLLNGEEKEVDTSKIRDVSRQGKNVKFLYNDNGKYGAGNVDEKDAPKELLWML

WDAEANKWLSLLHSLIEEAQNEREKNEQELL

Sac7d-Cpep2.3

VKVKFLLNGEEKEVDTSKIRDVSRQGKNVKFLYNDNGKYGAGNVDEKDAPKELLWML

WDAEANKWLSLLHSEAEERQNELEKNEQELL

Sac7d-Cpep2.4

VKVKFLLNGEEKEVDTSKIRDVSRQGKNVKFLYNDNGKYGAGNVDEKDAPKELLWML

WDAEANKWLSLLHSTIEEAQNLGEKNEQELL

Sac7d-Cpep2.5

VKVKFLLNGEEKEVDTSKIRDVSRQGKNVKFLYNDNGKYGAGNVDEKDAPKELLWML

WDAEANKWLSLLHSIVEELQNTLEKNEQELL

Sac7d-Cpep2.6

VKVKFLLNGEEKEVDTSKIRDVSRQGKNVKFLYNDNGKYGAGNVDEKDAPKELLWML

WDAEANKWLSLLHSILEEIQNAMEKNEQELL

Sac7d-Cpep2.7

VKVKFLLNGEEKEVDTSKIRDVSRQGKNVKFLYNDNGKYGAGNVDEKDAPKELLWML

WDAEANKWLSLLHSILEEAQNDREKNEQELL

Sac7d-Cpep2.8

VKVKFLLNGEEKEVDTSKIRDVSRQGKNVKFLYNDNGKYGAGNVDEKDAPKELLWML

WDAEANKWLSLLHSLFEELQNAQEKNEQELL

Sac7d-Cpep2.9

VKVKFLLNGEEKEVDTSKIRDVSRQGKNVKFLYNDNGKYGAGNVDEKDAPKELLWML

WDAEANKWLSLLHSMIEELQNVAEKNEQELL

Sac7d-Cpep2.10

VKVKFLLNGEEKEVDTSKIRDVSRQGKNVKFLYNDNGKYGAGNVDEKDAPKELLWML

WDAEANKWLSLLHSLLEEMQNQQEKNEQELL

Sac7d-Cpep2.11

VKVKFLLNGEEKEVDTSKIRDVSRQGKNVKFLYNDNGKYGAGNVDEKDAPKELLWML

WDAEANKWLSLLHSLMEEGQNLQEKNEQELL

Sac7d-Cpep2.12

VKVKFLLNGEEKEVDTSKIRDVSRQGKNVKFLYNDNGKYGAGNVDEKDAPKELLWML

WDAEANKWLSLLHSALEEAQNQGEKNEQELL

Sac7d-Cpep2.13

VKVKFLLNGEEKEVDTSKIRDVSRQGKNVKFLYNDNGKYGAGNVDEKDAPKELLWML

WDAEANKWLSLLHSLIEEGQNMSEKNEQELL

Sac7d-Cpep2.14

VKVKFLLNGEEKEVDTSKIRDVSRQGKNVKFLYNDNGKYGAGNVDEKDAPKELLWML

WDAEANKWLSLLHSILEEIQNEIEKNEQELL

Sac7d-Cpep2.15

VKVKFLLNGEEKEVDTSKIRDVSRQGKNVKFLYNDNGKYGAGNVDEKDAPKELLWML

WDAEANKWLSLLHSLVEEGQNIQEKNEQELL

Sac7d-Cpep2.16

VKVKFLLNGEEKEVDTSKIRDVSRQGKNVKFLYNDNGKYGAGNVDEKDAPKELLWML

WDAEANKWLSLLHSIIEELQNLNEKNEQELL

Sac7d-Cpep2.17

VKVKFLLNGEEKEVDTSKIRDVSRQGKNVKFLYNDNGKYGAGNVDEKDAPKELLWML

WDAEANKWLSLLHSAIEEAQNLVEKNEQELL

Sac7d-Cpep2.18

VKVKFLLNGEEKEVDTSKIRDVSRQGKNVKFLYNDNGKYGAGNVDEKDAPKELLWML

WDAEANKWLSLLHSIVEEAQNWAEKNEQELL

Sac7d-Cpep2.19

VKVKFLLNGEEKEVDTSKIRDVSRQGKNVKFLYNDNGKYGAGNVDEKDAPKELLWML

WDAEANKWLSLLHSLLEEAQNRHEKNEQELL

Sac7d-Cpep2.20

VKVKFLLNGEEKEVDTSKIRDVSRQGKNVKFLYNDNGKYGAGNVDEKDAPKELLWML

WDAEANKWLSLLHSFLEEAQNSQEKNEQELL

REFERENCES

1. Riddell, J., K.R. Amico, and K.H. Mayer, *HIV Preexposure Prophylaxis: A Review.* JAMA, 2018. **319**(12): p. 1261-1268.
2. Cihlar, T. and M. Fordyce, *Current status and prospects of HIV treatment.* Current Opinion in Virology, 2016. **18**: p. 50-56.
3. Brandenberg, O.F., et al., *The HIV-1 Entry Process: A Stoichiometric View.* Trends in Microbiology, 2015. **23**(12): p. 763-774.
4. Davenport, Y.W., A.P. West, and P.J. Bjorkman, *Structure of an HIV-2 gp120 in Complex with CD4.* Journal of virology, 2016. **90**(4): p. 2112.
5. Weiss, C.D., *HIV-1 gp41: mediator of fusion and target for inhibition.* AIDS reviews, 2003. **5**(4): p. 214.
6. Jason, L., L. John, and K. Peter, *Enfuvirtide.* Nature Reviews Drug Discovery, 2003. **2**(5): p. 345.
7. Tom, M., et al., *Enfuvirtide: the first therapy to inhibit the entry of HIV-1 into host CD4 lymphocytes.* Nature Reviews Drug Discovery, 2004. **3**(3): p. 215.
8. Patel, I., et al., *Pharmacokinetics, Pharmacodynamics and Drug Interaction Potential of Enfuvirtide.* Clinical Pharmacokinetics, 2005. **44**(2): p. 175-186.
9. Huet, T., et al., *Long-lasting enfuvirtide carrier pentasaccharide conjugates with potent anti-human immunodeficiency virus type 1 activity.* Antimicrobial agents and chemotherapy, 2010. **54**(1): p. 134.

10. Wei, X., et al., *Emergence of Resistant Human Immunodeficiency Virus Type 1 in Patients Receiving Fusion Inhibitor (T-20) Monotherapy.* Antimicrobial Agents and Chemotherapy, 2002. **46**(6): p. 1896.

11. Xu, L., et al., *Emergence and Evolution of Enfuvirtide Resistance following Long-Term Therapy Involves Heptad Repeat 2 Mutations within gp41.* Antimicrobial Agents and Chemotherapy, 2005. **49**(3): p. 1113.

12. Welch, B.D., et al., *Design of a Potent D-Peptide HIV-1 Entry Inhibitor with a Strong Barrier to Resistance.* The Journal of Virology, 2010. **84**(21): p. 11235.

13. Horne, W.S., et al., *Structural and biological mimicry of protein surface recognition by α/β-peptide foldamers.* Proceedings of the National Academy of Sciences, 2009. **106**(35): p. 14751.

14. Samuel, K.S. and S.K. Peter, *Protein grafting of an HIV-1-inhibiting epitope.* Proceedings of the National Academy of Sciences of the United States of America, 2003. **100**(17): p. 9756.

15. Tennyson, R.L., et al., *Helix-Grafted Pleckstrin Homology Domains Suppress HIV-1 Infection of CD4-Positive Cells.* ChemBioChem, 2016. **17**(20): p. 1945-1950.

16. Walker, S.N., et al., *GLUE That Sticks to HIV: A Helix-Grafted GLUE Protein That Selectively Binds the HIV gp41 N-Terminal Helical Region.* ChemBioChem, 2015. **16**(2): p. 219-222.

17. Lemmon, M.A., *Pleckstrin homology domains: not just for phosphoinositides.* Biochemical Society Transactions, 2004. **32**: p. 707-711.

18. Lemmon, M.A., *Pleckstrin homology (PH) domains and phosphoinositides.* Cell Biology Of Inositol Lipids And Phosphates, 2007. **74**: p. 81-93.

19. Lemmon, M.A. and K.M. Ferguson, *Pleckstrin homology domains.* Current topics in microbiology and immunology, 1998. **228**: p. 39.

20. Root, M.J., M.S. Kay, and P.S. Kim, *Protein Design of an HIV-1 Entry Inhibitor.* Science, 2001. **291**(5505): p. 884-888.

21. Tennyson, R.L., et al., *Evaluation of sequence variability in HIV-1 gp41 C-peptide helix-grafted proteins.* Bioorganic & Medicinal Chemistry, 2018. **26**(6): p. 1220-1224.

22. Behar, G., et al., *Tolerance of the archaeal Sac7d scaffold protein to alternative library designs: characterization of anti-immunoglobulin G Affitins.* Protein Engineering, Design & Selection, 2013. **26**(4): p. 267-275.

23. Ikeda, T., Tennyson, R.L., Walker, S.N., Harris, R.S., McNaughton, B.R., *Evolved Biologics Designed for Monthly Administration Potently Inhibit Entry of Enfuvirtide-Resistant HIV-1.* ACS Infectious Disease, 2019.

24. Haché, G., et al., *Evolution of HIV-1 Isolates that Use a Novel Vif-Independent Mechanism to Resist Restriction by Human APOBEC3G.* Current Biology, 2008. **18**(11): p. 819-824.

25. Boder, E. and K. Wittrup, *Yeast surface display for directed evolution of protein expression, affinity, and stability*, in *Methods Enzymol.* 2000. p. 430-444.

26. Yuxian, H., et al., *Potent HIV fusion inhibitors against Enfuvirtide-resistant HIV-1 strains.* Proceedings of the National Academy of Sciences, 2008. **105**(42): p. 16332.

27. Vaneycken, I., et al., *In vitro analysis and in vivo tumor targeting of a humanized, grafted nanobody in mice using pinhole SPECT/micro-CT.* J Nucl Med, 2010. **51**(7): p. 1099-1106.

28. Wittrup, K., et al., *Practical Theoretic Guidance for the Design of Tumor-Targeting Agents.* Methods In Enzymology: Protein Engineering For Therapeutics, Vol 203, Pt B, 2012. **503**: p. 255-268.

29. Schmidt, Michael m., et al., *Crystal Structure of an HSA/FcRn Complex Reveals Recycling by Competitive Mimicry of HSA Ligands at a pH-Dependent Hydrophobic Interface.* Structure, 2013. **21**(11): p. 1966-1978.

30. Wu, B. and Y.N. Sun, *Pharmacokinetics of Peptide–Fc Fusion Proteins.* Journal of Pharmaceutical Sciences, 2014. **103**(1): p. 53-64.

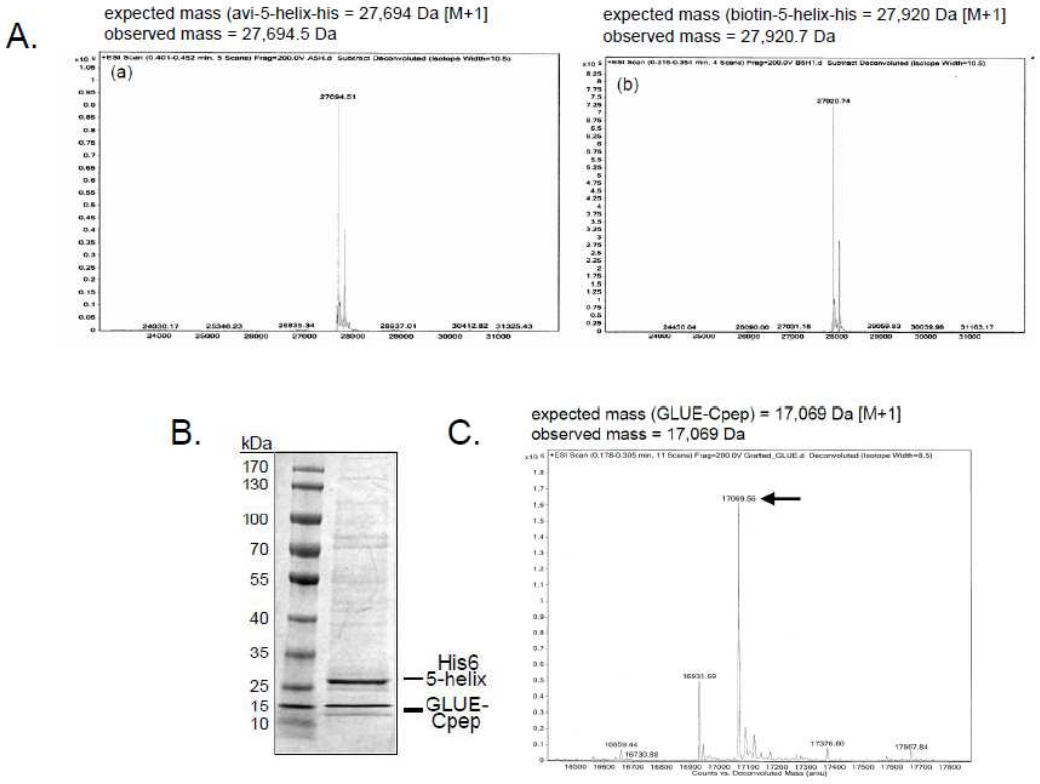
31. Chia, J., et al., *Intracellular trafficking and FcRn-dependent recycling of recombinant factor VIIa-albumin fusion protein (rVIIa-FP) provides a mechanism for half-life extension in vivo.* Haemophilia, 2016. **22**: p. 77-77.

32. Dennis, M.S., et al., *Albumin binding as a general strategy for improving the pharmacokinetics of proteins.* Journal Of Biological Chemistry, 2002. **277**(38): p. 35035-35043.

33. Zhao, H.L., et al., *Increasing the homogeneity, stability and activity of human serum albumin and interferon-α2b fusion protein by linker engineering.* Protein Expression and Purification, 2008. **61**(1): p. 73-77.

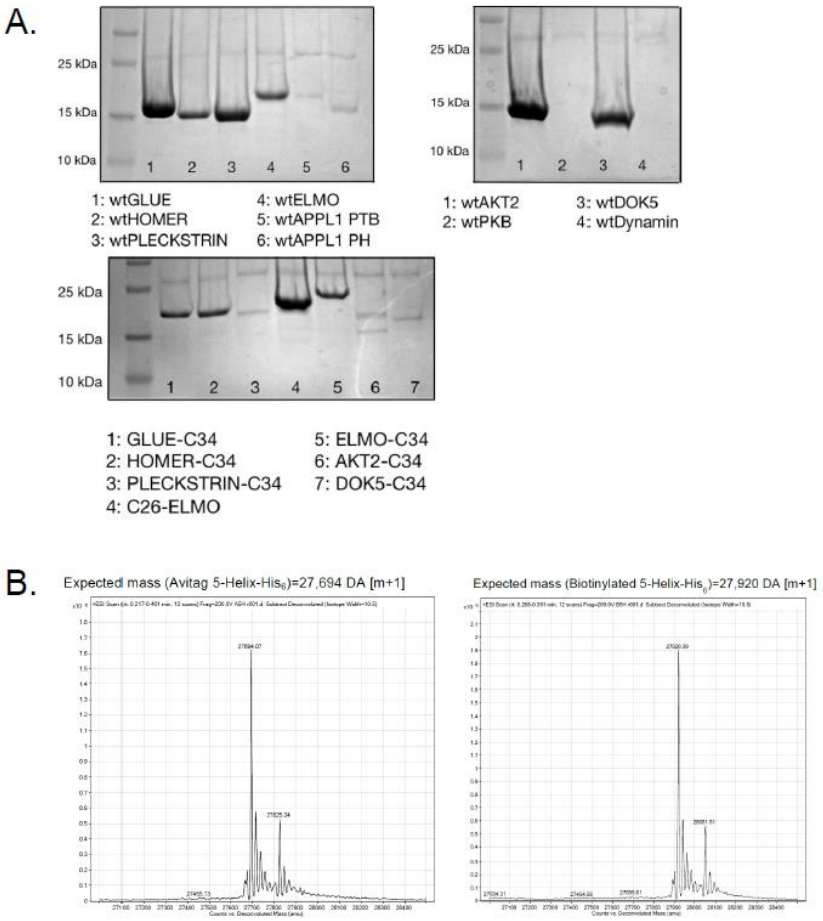
APPENDICES

6.1 Chapter 2 Supplemental Data



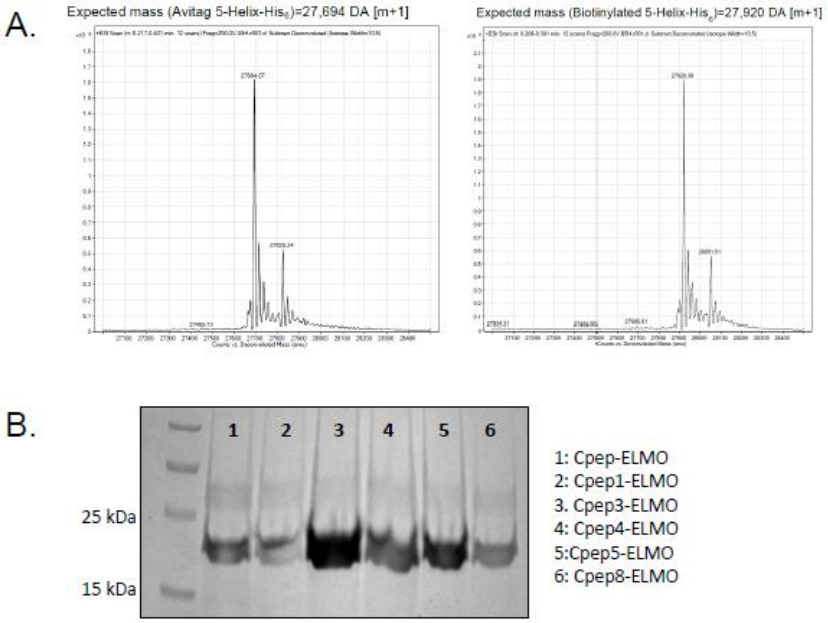
Supplemental Figure 6.1 (A) Mass spectrometry of Avi-tagged 5-helix-his and biotinylated avi-5-helix-his shows 100% biotin conjugation. (B) Nickel pulldown of *E. coli* cell lysate expressing 5-helix-his and untagged GLUE-Cpep in a 1:1 ratio (densitometry data not shown). (C) Mass spectrometry of *E. coli* cell lysate expressing 5-helix-his and untagged GLUE-Cpep after Nickel NTA purification. Untagged GLUE-Cpep is copurified with 5-helix-his indicating binding in complex cellular milieu.

6.2 Chapter 3 Supplemental Data



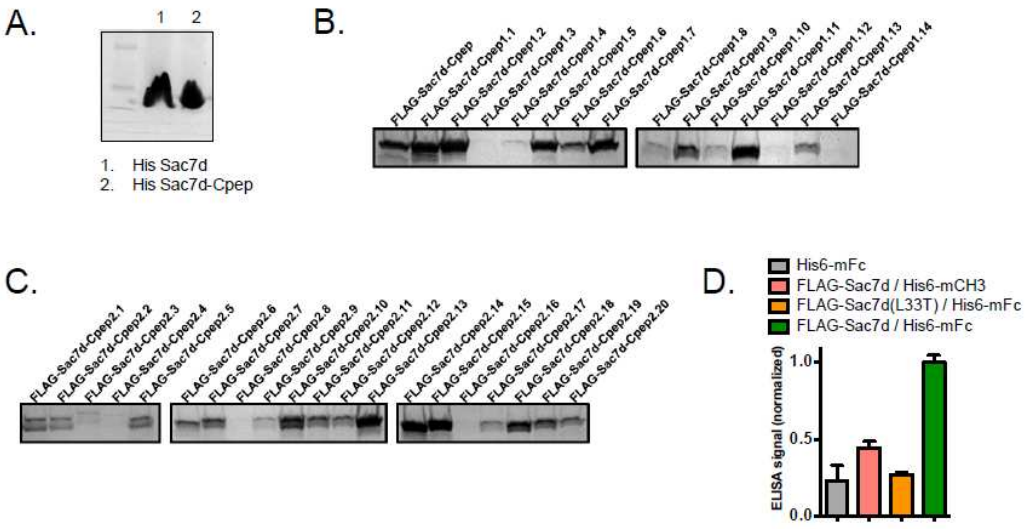
Supplemental Figure 6.2 (A) Protein expression of native scaffold candidates through Ni-NTA column purification and analysis via SDS-PAGE. (B) Mass spectrometry data of biotinylation reaction displaying 100% conjugation from Avi-tagged 5-helix-his (left, 27,694 Da) to biotin-5-helix-his (right, 27,920 Da).

6.3 Chapter 4 Supplemental Data



Supplemental Figure 6.3 A) Mass spectrometry of Avi-tagged 5-helix-his (left) and biotinylated avi-5-helix-his (right) shows 100% biotin conjugation to protein. (B) SDS-PAGE confirmation of Cpep-ELMO and Cpep-ELMO mutants used in this work (Cpep1-ELMO, Cpep3-ELMO, Cpep4-ELMO, Cpep5-ELMO, and Cpep8-ELMO).

6.4 Chapter 5 Supplemental Data



Supplemental Figure 5.11 (A) SDS-PAGE gel displaying protein expression and purity of Sac7d scaffold (lane 1) and helix-grafted Sac7d-Cpep (lane 2). (B) SDS-PAGE gel displaying protein expression of FLAG-tagged Sac7d-Cpep and generational variants. (C) *In vitro* ELISA of binding interaction between Sac7d and either mCH3 or mFC. Sac7d L33T mutation reduces ELISA signal (orange bar) thus turning off binding to mFc. Error bars represent the standard error of the mean (SEM) for three separate experiments. If error bars are not visible, error is smaller than the data point.

Supplemental Table 5.11 Library screening schematic for all 5 rounds of cell sorting.

	[biotin-5-helix, nM]	[Cpep-ELMO competitor, nM]	[Sac7d-Cpep competitor, nM]	Yeast Cells Screened	Yeast Cells Sorted	% double positive
Generation 1: Round 1	1.00	10.00	0.00	~7×10^7	~5.25×10^5	0.75%
Generation 1: Round 2	0.10	10.00	0.00	~1×10^7	~1.0×10^6	1.00%
Generation 2: Round 1	4.00	0.00	0.00	~8×10^7	~2.4×10^6	3.00%
Generation 2: Round 2	0.20	0.00	0.00	~2×10^7	~5.0×10^5	2.50%
Generation 2: Round 3	0.05	0.00	1.00	~6×10^6	~3.0×10^4	0.05%