DISSERTATION

Protein Engineering Strategy for the Stabilization of HIV-1 α-Helical Peptides

Submitted by

Rachel Lee Tennyson

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

Christopher Ackerson
Christopher Snow
Daniel Gustafson

ABSTRACT

Protein Engineering Strategy for the Stabilization of HIV-1 α-Helical Peptides

Many disease-relevant protein-protein interactions (PPIs) contain an alpha helix and helical binding cleft at their interface. Disruption of these interactions with helical peptide mimics is a validated therapeutic strategy. However, short peptides typically do not fold into stable helices, which significantly lowers their *in vivo* stability. Researches have reported methods for helical peptide stabilization but, these approaches rely on laborious, and often expensive, chemical synthesis and purification. The research I have preformed aims to stabilize disease-relevant helices through protein engineering. In contrast to chemically constrained helical peptides, a protein can be expressed in a cellular system on a much larger scale. Recently, we reported a new strategy termed "helix-grafted display" that overcomes the traditional hurdles of helical mimics and applied it to the challenge of suppressing HIV entry. Our helix grafted proteins, potently inhibits formation of the extracellular PPI involving C-peptide helix, and HIV gp41 N-peptide trimer, as tested in HIV CD4+ cells. Further optimization of the helical sequence by yeast display yielded new proteins that suppress HIV-1 entry and express substantially better in *E. coli*. Furthermore, fusion proteins designed to improve the serum stability of these helix grafted proteins have been made that potently suppress HIV-1 entry. Collectively, I report a potential cocktail of evolved HIV-1 entry inhibitors that are functional against an Enfuvirtide-resistant strain and are designed for serum stabilities that rival current monoclonal antibody drugs.

ACKNOWLEDGEMENTS

First and foremost I would like to start off by thanking my incredible family. There wasn't a time you weren't willing to listen to everything I had to say. I am here because of you. Thank you for teaching me to keep trying. I would be lost without your guidance and support. Mom, thank you for always being my cheerleader and best friend. Dad, thank you for always being my pillar of support, you taught me how to keep fighting for my goals. Nola, I'm incredibly grateful for all your amazing advice through school and for proofreading all my documents. To my grandparents and aunt, I love you. The love you have is unconditional and I am forever grateful.

To my close friends, I can't thank you enough for your support. Thank you for making me laugh when times seemed tough and for making me smile when things seemed grim. I hope I can do the same for you.

Through my time in graduate school I had two advisors, Brian McNaughton and Alan Kennan, who I would like to thank individually. Brian, thank you for teaching me to be ambitious through school and my career. From the beginning you told me if I worked hard I could make it to where I want to be. Thank you for thinking about the next steps, for pushing for publications and for always applying for funding. Alan, thank you for always having your door open. I came to you with every question and you were always supportive. Thank you for always encouraging me. My advisors taught me how to be a better graduate student and for that I am forever grateful.

I would like to thank my peers in the McNaughton lab. You have all taught me something different, for which I am grateful. To my coworker Susanne Walker, I've enjoyed seeing the

progression of the project. We got to see the beginning and end together and I'm so grateful to have you as a labmate and friend. To Alex Chapman, thank you for being my mentor. Thank you for all you input, whether it was project related, career related, or outside of work. Your feedback has been and continues to be important to me.

I would also like to thank our collaborators at the University of Minnesota. I never got to formally meet Teru Ikeda or Reuben Harris but I've had the pleasure of working with them the past three years. I would also like to express my gratitude to my committee members. Thank you for good scientific conversations for feedback on my dissertation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

Human Immunodeficiency Virus, Drugs and Targeting Membrane Fusion

1.1 Impacts of Human Immunodeficiency Virus and Acquired Immune Deficiency Syndrome

Human Immunodeficiency Virus (HIV) is considered one of the most dangerous viruses known to mankind. HIV is able to infect a host's healthy immune cell, integrate its viral DNA into the human genome, and hijack the cells machinery to produce more virions.[1] While the mechanism is not entirely understood, the once healthy (immune) T-cells die and a person's immune system slowly deteriorates. When the T-cell count reaches below 200 (normal T-cell count is 500-1500) patients are diagnosed with Acquired Immune Deficiency Syndrome (AIDS).[2] HIV/AIDS has killed approximately 35 million due to AIDS related illnesses and there are currently 37 million people living with HIV/AIDS as of 2017.[3, 4] Reports project there will be 2 million newly infected HIV+ patients every year and an estimated one million will die worldwide despite having medications that can prevent HIV infection and therapeutics that can slow down the progression of AIDS.[1]

It is believed that HIV first emerged in West Africa during the 1930s, most likely during a hunting expedition when a man came into contact with blood from a chimpanzee.[5] The first five reported cases of HIV in the United States were reported on June 5 1981.[6] By 1990 an estimated 10 million people were living with HIV worldwide. This number jumped to 30 million in just seven years and by 1999, 14 million people died due to AIDS related illnesses since the start of the viral outbreak.

With such a high mortality rate in a short matter of time, HIV+ patients and the world began searching for therapeutics to treat the virus and prevent infection. Six years after the initial reported cases of AIDS in the United States, the FDA approved its first small molecule HIV drug. However, HIV has been able to evade therapeutic intervention due to its adaptability. The virus is able to replicate 100,000 billion times per day, which allows it to continuously attack healthy cells. The virus also has a reverse transcriptase (discussed further below) that is highly prone to inserting mutations in the viral genome, allowing the virus to evolve quickly. HIV's high replication rate and ability to mutate and evolve resulted in a drug resistance problem. In 1996 it was discovered that patients needed to be enrolled in antiretroviral therapy.[7] Antiretroviral therapy, also known as ART, is a combination therapy of at least two drugs that target different parts of viral infection. Patients who are prescribed ART must follow the strict regime of drugs for the rest of their lives to combat the virus' ability to mutate and evolve. Combination therapy put a large demand on the number of available drugs targeting HIV. In some cases, patients were taking an astounding 6 pills a day and paying $10,000 per year, but those patients were able to combat the progression of AIDS. Despite having 28 FDA approved drugs in 2018, many patients throughout the world cannot gain access to these treatments because of their enormous cost.

1.2 HIV-1 Viral Life Cycle

To build a large repertoire of drugs targeting HIV, researchers needed to understand how the virus works to tailor drugs to interrupt specific protein interactions. HIV is a retrovirus meaning it's a type of RNA virus.[8] HIV first enters the bloodstream and binds to a CD4+ cell (healthy immune/T-cell, Step 1). The viral membrane then undergoes conformational rearrangements

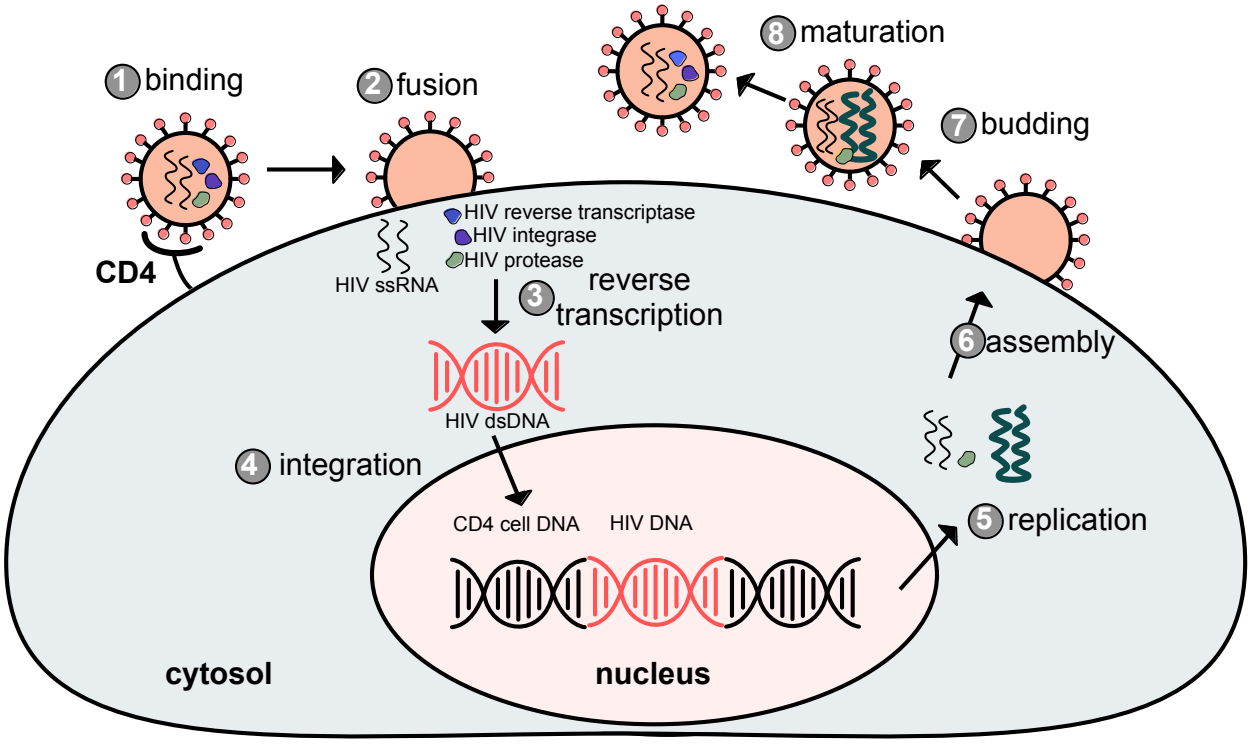


Figure 1.1 Schematic of HIV viral life cycle. The virus recognizes CD4+ immune cell, injects its RNA and viral proteins. The virus is able to hijack the hosts cells machinery to integrate its viral DNA and produce more virions.

with the host cell allowing for the virus to inject its genetic material (Step 2). This genetic material includes its single stranded RNA (ssRNA), and its enzymes including reverse transcriptase, integrase and protease. The ssRNA is then transcribed into double stranded DNA by HIV's reverse transcriptase enzyme (Step 3). The DNA is then shuttled into the host cells nucleus along with integrase. The host cells DNA is then cut and HIV's double stranded DNA is inserted, or integrated into the genome (Step 4). From there the virus is able to use the host cells machinery to turn on transcription and translation to produce long chains of viral protein (Step 5). This long protein chain is then transported towards the surface of the cell. These proteins then assemble into an immature virion (Step 6). The immature HIV particle begins to bud off (Step 7) and the long protein chain is cut by HIV's protease to produce smaller functional proteins to make a mature virion (Step 8). These virions can then move on to attack a new CD4+ cell.[9]

1.3 Current Treatments for HIV-1

There are currently 28 individual drugs on the market targeting HIV and most are taken in combination as an antiretroviral therapy. The drugs are classified under 4 categories based on their targets: viral entry, reverse transcriptase (non-nucleoside and nucleoside), integrase or protease. Most of the current drugs on the market are small molecule inhibitors targeting HIV's reverse transcriptase or protease. Surprisingly, of the 28 approved drugs only three of them target viral entry, and two of them are biologics (peptide or antibody drugs). Viral entry inhibitors are able to block initial receptor binding, co-receptor activation or membrane fusion. While these viral entry inhibitors are effective, they are considered a last resort treatment because they are costly to make, require multiple injections and/or have severe side effects. Due to its rapid replication cycle, there is a strong need for improved entry inhibitors that can be produced on a large scale and remain in the blood stream. The focus of this thesis pertains to investigating a protein engineering solution to inhibit fusion of the HIV viral particle to healthy immune cells. This therapeutic needs to be produced on a large scale, display good *in vivo* stability, be readily evolvable and be a potent membrane fusion inhibitor

1.4 Membrane Fusion and Entry Inhibitors

HIV uses its glycoproteins to recognize CD4 receptors on the outside of T-cells. Gp120 is what recognizes and binds CD4 on the host cell. Gp120 then sheds itself allowing gp41 to insert a N-terminal fusion peptide into the target cells bilayer, forming a bridge between the virus membrane and target cell membrane, known as the pre-hairpin intermediate. The pre-hairpin intermediate consists of a trimer of hairpins that expose both N and C peptides. After the virus anchors itself to the immune cell it folds over and brings the two membranes closer together

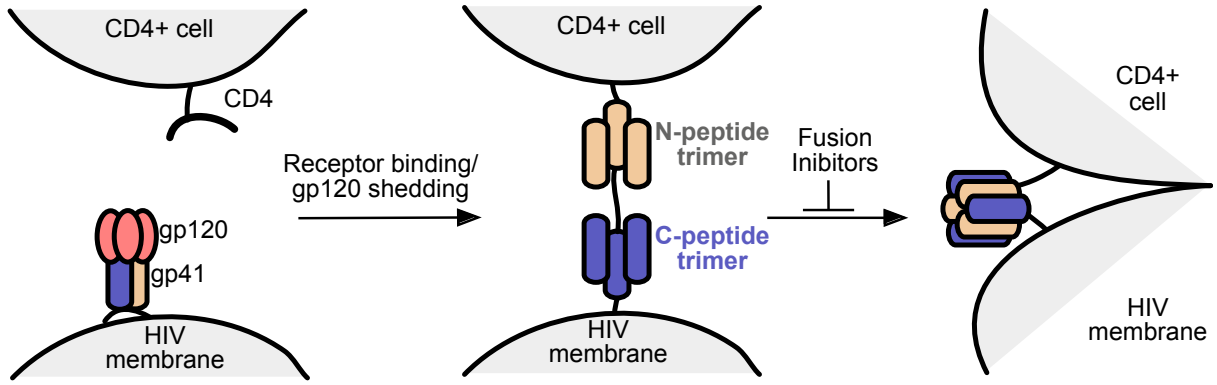


Figure 1.2 HIV-1 membrane fusion is enabled by HIV's glycoproteins. GP120 recognizes the CD4 receptor on immune cells. Gp120 undergoes a confirmation rearrangement exposing gp41. GP41 is able to anchor itself into the host cell's membrane leading to a 'trimer-of-hairpins' assembly involving N-peptide and C-peptide trimers. This leads to juxtaposition of viral and immune cell membranes, and viral entry.

forming a six helical bundle. This bundle formation allows the virus to inject its genetic material ultimately leading to infection.

α-Helical peptides derived from the N-peptide trimer or C-peptide trimer of gp41 are a validated therapeutic approach for targeting membrane fusion. These peptides are able to bind to the exposed groves in the pre-hairpin intermediate and impede six helical bundle formation. NHR peptides are able to inhibit infection with micromolar concentrations but often aggregate because of their hydrophobic residues.[10] C-peptides have proven to be more potent and effective than NHR peptides and can effectively inhibit HIV-1 viral membrane fusion in nanomolar concentrations.[11] A crystal structure of the full length N- (tan) and C- peptide (navy) hairpin interaction is shown below (**Figure 1.3A).** Enfuvirtide (also known as T-20 or Fuzeon), is currently the only FDA approved peptide therapeutic for HIV that specifically blocks viral membrane fusion by binding the N peptide trimers and obstructing 6 helix bundle formation. Its 36 amino acid sequence directly comes from gp41. Making a peptide drug of such a long sequence requires laborious and expensive peptide synthesis.

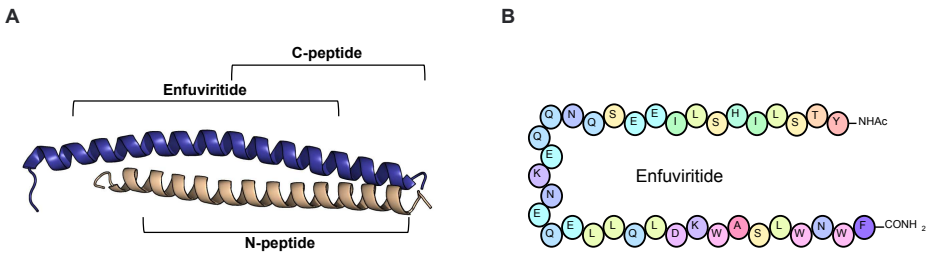


Figure 1.3 (*A*) GP41 derived membrane fusion inhibitors. (*B*) Amino acid sequence of Enfuvirtide

1.5 Drawbacks of α-Helical Peptides

Despite their success, α-helical peptides still have considerable drawbacks, namely cost of production and sensitivity to degradation. Enfuvirtide is a classic example of these setbacks and is considered to be a last resort treatment for HIV patients. Enfuvitide has a half of ~3.4 hours which requires patients to undergo twice daily injections which is less than ideal for them. Proteolytic stability is a common issue for α-helical peptides because they are often found disordered in solution, which makes them susceptible to degradation. In order for a peptide to bind its desired target a large entropic cost is required for it go from its partially unfolded state to an ordered conformation.[12] Peptide therapeutics have sought to overcome the limitations of helical peptides by methods of stabilization and structured mimics.[13] Helix stabilization constrains the peptide so it can no longer move freely. Common methods for this are forming salt bridges, chelation with metals, hydrocarbon stapling and covalent cyclization.[13] Helical mimics imitate the topography of the peptide, allowing for correct orientations of the functional groups. Approaches for this includes using α/β peptides, terphenyls, and peptoids.[14] These methods have shortcomings of their own and can result in a decrease of the original peptide's potency.[15] Increased optimization is still of significant interest despite the development of new strategies for cost effective and efficient peptide synthesis.

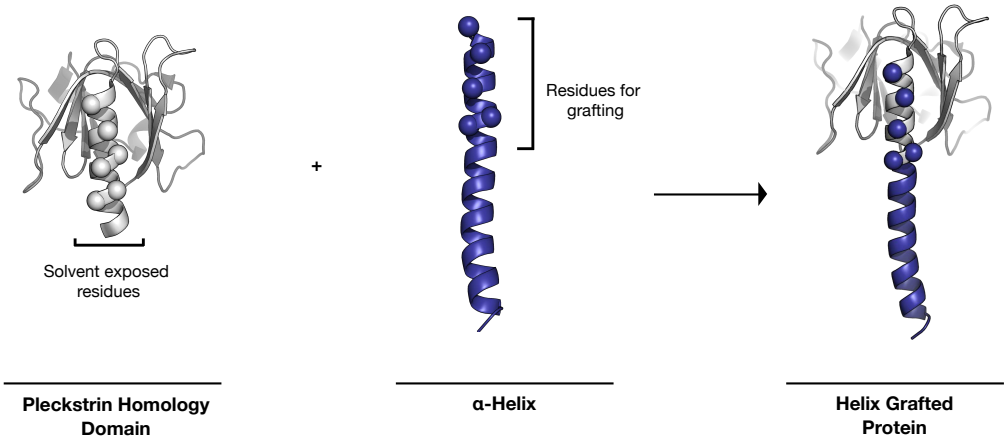


Figure 1.4 Proposed protein engineering strategy for creating stable HIV membrane fusion inhibitors. Solvent exposed residue on a PH domain we will be mutated to match residues from a HIV-1 disease relevant peptide. The resulting helix-grafted protein will be expressed in *E. coli* and tested for function.

1.6 Protein Engineering Solution to Helix Stabilization

Protein grafting could be an ideal method for helix stabilization due to its large success with N- and C-peptides as HIV-1 gp41-1 binders. Protein grafting is used to transfer the biological function of a ligand onto the surface of another protein.[16] Kim and co-workers applied this method with the C-peptide of HIV-1, grafting it onto a GCN4 leucine zipper.[15] Their protein, C34-GCN4, showed similar activity to the native C-peptide. However, helix stabilization is still a challenge when grafting onto a pre-established helical interface and still required expensive peptide synthesis. We hypothesized a protein with a solvent exposed helix could be stabilized by its tertiary structure and would be better suited for grafting since it can be expressed in *E coli*.

Pleckstrin Homology (PH) domains have a C-terminal solvent exposed helix that fits in a cleft formed by the surrounding β-sheets.[17] These domains are generally found in proteins involved in cell trafficking and function in binding phosphoinositides.[13] These proteins are ~120 residues which makes for quick insertion into a bacterial vector. The native function of PH domains can also be turned off by a single mutation via site directed mutagenesis. The hydrophobic effect suggests that the buried residues are critical for providing the protein's

tertiary structure. If those residues were untouched and only solvent exposed helical residues mutated, the interface could be altered to match the binding face of the peptide. **Figure 1.4** depicts the helix grafted display method developed in the McNaughton/Kennan laboratories

1.7 Scope of This Thesis

This thesis will describe a protein engineering solution to stabilizing an HIV relevant peptide, termed C-peptide. By using PH domains as a generic scaffold we were able to mutate and extend their helix to mimic that of C-peptide. These proteins could be produced in *E. coli* and remained folded in solution, overcoming proteolytic stability issues and laborious chemical synthesis associated with α-helical peptides. These proteins tightly bind a model of membrane fusion and are also potent fusion inhibitors in a live virus assay. We also found that the solvent exposed helix was amenable to mutations and could be optimized for higher binding affinity to a membrane fusion model. By diversifying the helix we found numerous fusion inhibitors that could potentially be used as a cocktail therapy. Our helix-grafted proteins can also be fused to antibody mimics for stability, giving these proteins the potential to remain in the bloodstream for long periods of time and still inhibit HIV viral infection.

Chapter II details our initial helix grafted display efforts on a scaffold called GLUE. GLUE, (GRAM-Like Ubiquitin-binding in EAP45) is a PH domain endogenous to yeast, which can be easily expressed in *E. coli* with a His6 tag. We found that GLUE's native helix can be mutated to match that of HIV's gp41 C-peptide. We termed this engineered protein GLUE Cpep. GLUE Cpep is folded in solution, stable in human serum, and soluble in aqueous solutions and thus overcomes challenges faced by peptides derived from HIV gp41. [18]

Chapter III is an expansion of our proof concept model, GLUE-Cpep. To develop the helix grafted display platform we prepared a number of human-derived helix grafted PH domains of varied helix length and measured properties relevant to therapeutic and basic research applications. In particular we showed that some of these new reagents expressed well as recombinant proteins, were relatively stable in human serum, bound a mimic of pre-fusogenic HIV-1 gp41 *in vitro* and in complex biological environments, and significantly lowered the incidence of HIV-1 infection of CD4-positive cells. [19]

In Chapter IV we take Cpep ELMO, a helix grafted protein found in Chapter III and used yeast display to screen a library of grafted C-peptide helices for N-peptide trimer recognition. Using hits from the yeast display library screening, we evaluated the effect the helix mutations have on structure, expression, stability function and suppression of HIV entry in a live virus assay.[20]

Chapter IV describes our most potent entry inhibitor yet, Sac7d-Cpep. Our previous scaffolds in Chapter III and IV inhibited membrane fusion in a live virus assay, albeit, modestly (IC_{50} 190 nM - >1 μM). Sac7d Cpep on the other hand is able to suppress HIV-1 infection with an IC_{50} of 1.9-12.4 nM. This is competitive with Enfuvirtide and C-peptide, whose IC50s are 0.7-3.5 nM and 1.0-6.7 nM, respectively. Yeast display sequence optimization of solvent exposed helix residues, similar to the work done in Chapter IV, revealed new proteins with improved expression in *E. coli*, and no appreciable change in entry inhibition. Evolved proteins inhibit the entry of a clinically relevant mutant of HIV-1 that is gp41 C-peptide sensitive and Enfuvirtide-resistant. Fusion proteins were then designed for serum stability and also potently suppress HIV-1 entry. [21]

This thesis describes four different stories entailing the development and evolution of stable HIV-1 gp41 C-peptide membrane fusion inhibitors. We found that the starting PH domain or PH-like domain scaffold needs to be robustly expressed in *E. coli* and amenable to mutation. Since viral membrane fusion is an enclosed space the inhibitors potency is size dependent. Some of our newly found helix grafted proteins are able to inhibit HIV infection in Enfuvirtide resitant strains, which goes to show the adaptability and utility of our helix-grafted display platform. These new biologics can be fused to antibody-derived proteins for improved *in vivo* stability.

REFERENCES

1. Ashkenazi, A.; Wexler-Cohen, Y.; Shai, Y., Multifaceted action of Fuzeon as virus-cell membrane fusion inhibitor. *Biochim Biophys Acta* **2011,** *1808* (10), 2352-8.

2. Hughson, G. Factsheet CD4 cell counts. http://www.aidsmap.com/CD4-cell-counts/page/1044596/ (accessed Feb 15, 2018).

3. Justman, J. E.; Mugurungi, O.; El-Sadr, W. M., HIV Population Surveys - Bringing Precision to the Global Response. *N Engl J Med* **2018,** *378* (20), 1859-1861.

4. UNAIDS Fact Sheet- World AIDS Day 2018. http://www.unaids.org/sites/default/files/media_asset/UNAIDS_FactSheet_en.pdf (accessed March 4, 2018).

5. Research, C. F. f. A. HIV and AIDS History. (accessed February 15, 2018).

6. AIDS.gov A Timeline of HIV and AIDS. (accessed February 15, 2018).

7. Vella, S.; Schwartlander, B.; Sow, S. P.; Eholie, S. P.; Murphy, R. L., The history of antiretroviral therapy and of its implementation in resource-limited areas of the world. *AIDS* **2012,** *26* (10), 1231-41.

8. Weiss, R. A., Retrovirus classification and cell interactions. *J Antimicrob Chemother* **1996,** *37 Suppl B*, 1-11.

9. Sleasman, J. W.; Goodenow, M. M., 13. HIV-1 infection. *J Allergy Clin Immunol* **2003,** *111* (2 Suppl), S582-92.

10. Ashkenazi, A.; Shai, Y., Insights into the mechanism of HIV-1 envelope induced membrane fusion as revealed by its inhibitory peptides. *Eur Biophys J* **2011,** *40* (4), 349-57.

11. Eckert, D. M.; Kim, P. S., Mechanisms of viral membrane fusion and its inhibition. *Annu Rev Biochem* **2001,** *70*, 777-810.

12. Olmez, E. O.; Akbulut, B. S., *Protein-Peptide Interactions Revolutionize Drug Development*. 2012.

13. Azzarito, V.; Long, K.; Murphy, N. S.; Wilson, A. J., Inhibition of alpha-helix-mediated protein-protein interactions using designed molecules. *Nat Chem* **2013,** *5* (3), 161-73.

14. Guarracino, D. A.; Bullock, B. N.; Arora, P. S., Mini review: protein-protein interactions in transcription: a fertile ground for helix mimetics. *Biopolymers* **2011,** *95* (1), 1-7.

15. Sia, S. K.; Kim, P. S., Protein grafting of an HIV-1-inhibiting epitope. *Proc Natl Acad Sci U S A* **2003,** *100* (17), 9756-61.

16. Kritzer, J. A.; Zutshi, R.; Cheah, M.; Ran, F. A.; Webman, R.; Wongjirad, T. M.; Schepartz, A., Miniature protein inhibitors of the p53-hDM2 interaction. *Chembiochem* **2006,** *7* (1), 29-31.

17. Ferguson, K. M.; Kavran, J. M.; Sankaran, V. G.; Fournier, E.; Isakoff, S. J.; Skolnik, E. Y.; Lemmon, M. A., Structural basis for discrimination of 3-phosphoinositides by pleckstrin homology domains. *Mol Cell* **2000,** *6* (2), 373-84.

18. Walker, S. N.; Tennyson, R. L.; 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* (2), 219-22.

19. 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* (20), 1945-1950.

20. 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. *Bioorg Med Chem* **2018,** *26* (6), 1220-1224.

21. Tennyson, R. L.; Walker, S. N.; Ikeda, T.; Harris, R. S.; McNaughton, B. R., Evolved Proteins Inhibit Entry of Enfuviritde-Resistant HIV-1. *ACS Infect. Dis* **Just Accepted**.

CHAPTER TWO

GLUE That Sticks to HIV: A Helix-Grafted GLUE Protein That Potently and Selectively Binds HIV gp41 N-terminal Helical Region[1]

2.1 Attributions

In this work, led by fellow graduate student Susanne Walker, I assisted in molecular cloning, protein purification and ELISA experiments. I also performed all circular dichroism experiments.

2.2 Introduction

A medicinally significant subset of protein-protein interactions consists of those with α-helix bound to a surface cleft. This 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[1-3], covalently constrained (or 'stapled') peptides[4-7], and helical 'foldamers'[8-12] employing natural or unnatural backbone architectures (as briefly mentioned in Chapter I). Each has produced effective agents, but requires non-trivial synthetic effort and expense.

By exploiting the flexibility and feasibility of protein production in E. coli, its possible a more accessible suite of ligands could be generated. In developing a general scaffold for helix grafted display, we sought to identify a protein fold that presents 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

[1] This work has been adapted from:
Walker, S. N.; Tennyson, R. L.; 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* (2), 219-22

face. Ideally the basic scaffold would be readily expressible in soluble form, tolerant of mutation and/or extension of the helix, and protective against rapid helix degradation. The general concept of grafting is an established method for mimicking protein surfaces but has only been done by grafting one α-helix onto another.[13, 14]

Chapter II will describe our successful application of our specific 'helix-grafting' technique to a PPI crucial for HIV infection. We show that grafting of gp41 C-peptide residues onto the exposed helical face of a suitable host affords a new ligand that expresses well in *E. coli,* exhibits excellent serum stability and is capable of replicating the native interaction. Peptides derived from this C-peptide helix have been shown to bind the prehairpin intermediate and inhibit membrane fusion of HIV in human cells (as described in Chapter I).[15-17] We reasoned that a helix-grafted alternative might retain similar specificity but have improved stability, solubility, and availability.

2.3 GLUE as a Protein Scaffold for Helix Grafted Display

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 2A**, gray), which is derived from a subunit of the endosomal sorting complex.[18] Like other members of the PH family, the GLUE domain contains a C-terminal amphipathic helix resting in a cleft formed by two opposed beta 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

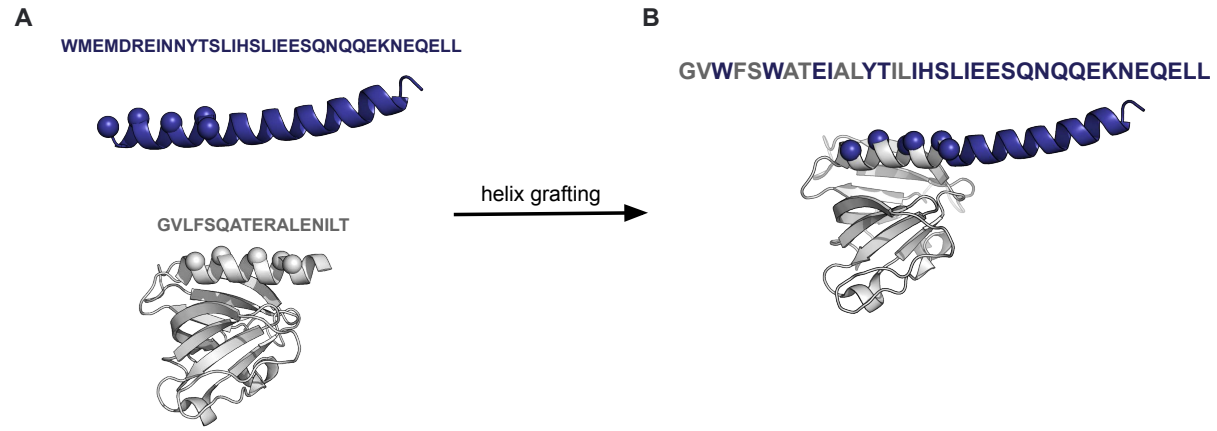


Figure 2.1 (*A*) Helix-grafting HIV gp41 C-peptide onto a stable PH domain protein. Wild-type GLUE (grey, bottom) and gp41 C-peptide (purple, top). Sequences of the GLUE helix and the corresponding responding region of the gp41 helix are shown in gray and purple, respectively. Spheres indicate the C-alpha position (*B*) Helix-grafted GLUE-Cpep, produced by backbone alignment of the independent structures. Spheres indicate C-alpha positions of GLUE residues mutated to those from gp41 (also coded in the sequence).

could be extended to a length comparable to the C-peptide 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.

We began by aligning the native helix on GLUE (**Figure 2.1A**, grey) with a single C-peptide helix from gp41 (**Figure 2.1A**, purple). Backbone atoms from the GLUE helix (PDB code 2CAY) were aligned with the corresponding number from the N-terminal segment of the gp41 C-peptide (PDB 1AIK), using PyMol's pair_fit algorithm. The overlay was very good (RMSD of 0.44 Å 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 attaching a pure gp41 sequence to the C-terminus of GLUE (**Figure 2.1B**) such that the total length of the new helix was appropriate for binding to the trimeric N-terminal coiled coil. The final sequence expressed as a soluble protein in *E. coli*.

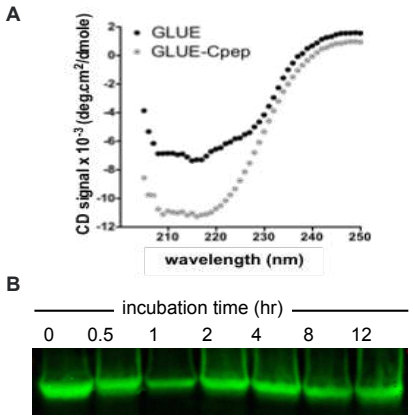


Figure 2.2 (*A*) Circular Dichroism (CD) data for wtGLUE (black) and GLUE-Cpep (gray). (*B*) Western blot of FLAG-tagged helix grafted GLUE that isn't incubated with human serum (lane 1), or after 0.5 to 12 hours of incubation (lanes 2-7).

2.4 Characterization of GLUE-Cpep

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 2.2A**). Both proteins display a similar overall signal, suggesting that the grafting process does not compromise the GLUE domain fold. Since 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. FLAG-tagged GLUE-Cpep incubated with human serum for up to 12 hours showed no appreciable degradation by Western blot analysis (**Figure 2.2B**). This supports a significant serum stability enhancement for the grafted protein compared to isolated peptides such as Enfuvirtide.

2.5 5-Helix Binding Model

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 coworkers. 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 sides already occupied, and just a single exposed interface (**Figure 2.3A**). Throughout, we use 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 2.3B**) exhibits a notably deeper signal for the 1:1 mixture than for either component alone, and the corresponding melt data (**Figure 2.3C**) reveal a dramatic increase in thermal stability, as evidenced by a significant shift of the overall melting curve, though the change in T_m is more modest (observed T_m values of ~77 °C, ~79 °C, ~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.

2.6 GLUE-Cpep Binds 5-Helix in Complex Cellular Environments

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.[19] *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-

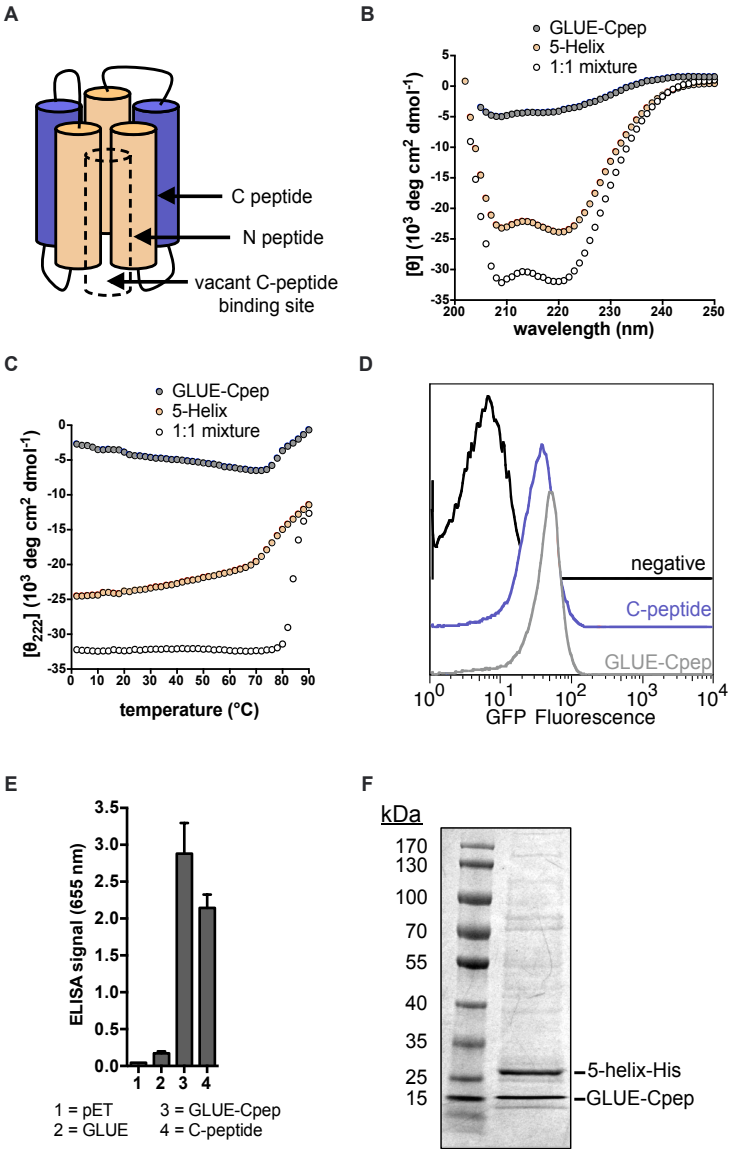


Figure 2.3 (*A*) Depection of 5-helix, a single protein consisting of three copies of gp41 N-peptide (tan) and two copies of gp41 C-peptide (purple). When folded, this protein presents a single binding site for a C-peptide (or mimic thereof), which is depicted as a dashed column. (*B*) Cirular dichroism spectra of GLUE-Cpep (blue circles), 5-helix (red circles) and a pre-mixed 1:1 ratio of 5-helix-His and GLUE-Cpep (open circles) (*C*) Circular dichroism data (222 nm) showing temperature-dependent melting of the solutions in part (*B*). (*D*) Flow cytometry data of *E. coli* following split-spGFP reassembly experiments. Negative = NscGFP-5-helix/CscGFP-; Cpep = NscGFP-5-helix/CscGFP-C-pep; GLUE-Cpep = NscGFP-5-helix/CscGFP-GLUE-Cpep. (*E*) ELISA data from *E. coli* cell lystate that contain an empty pET DUET plasmid, or pET DUET that encodes-5-helix-His along with wtGLUE, GLUE-Cpep, or C-peptide. (*F*) Co-purifciation of 5-helix-His and untagged GLUE-Cpep from *E. coli* cell lysate.

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 2.3D**). We further

characterized this interaction using an Enzyme-Linked Immunosorbant Assay (ELISA). The grafted GLUE binds 5-helix with slightly better affinity than the native C-peptide (**Figure 2.3E**, columns 3 and 4, respectively), while the wild type GLUE exhibits no appreciable affinity (**Figure 2.3E**, column 2), confirming the need for the grafted domain. This ELISA signal is observed even for a GLUE-Cpep sample that was pre-incubated with human serum (Supporting Information), confirming that the degradation-resistant form of the protein remains functional. Taken 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.

Binding selectivity was assessed by measuring the amount of protein that is co-purified from *E. coli* expressing an untagged GLUE-Cpep **(**~17.1 kDa) and 5-helix-His6 (~25.4 kDa). As seen in Figure 4F, the tagged 5-helix co-purifies with a single protein, which was identified as GLUE-Cpep by mass spectrometry (**Supplemental Data, 2.1**). The similar amounts of each co-purified protein (as determined by densitometry measurements of each protein band) further indicates that the complex involves a 1:1 ratio of proteins. The relatively miniscule levels of other co-purified cellular proteins indicates excellent selectivity for this interaction, even in a complex cellular environment, suggesting a reasonably strong mutual affinity.

2.7 Conclusion

We have demonstrated that the solvent exposed C-terminal alpha 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 co-purification 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 may represent a new reagent for inhibition of HIV entry. Additionally, helix-grafting onto PH and PH-like domains, such as GLUE, may 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.

2.8 Methods

Protein Expression and Purification Genes were cloned into pETDuet-1 using restriction enzymes BamHI and PacI, downstream of a His_6 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 (4950 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 shown below. Purified

protein concentrations were quantified using Beer's Law at an absorbance of 280nm, following standard practice. In general, GLUE-Cpep was expressed as a soluble protein (~3mg/L of *E. coli* culture).

Resolublization from Inclusion Bodies 5 Helix-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 5 Helix-His 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 (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. Purified proteins were quantified using Beer's Law at an absorbance of 280nm.

Circular Dichroism Proteins were purified as described above. Separately, each protein was diluted to 7-9 μM in Tris buffer (20 mM Tris pH 7.4, 100 mM NaCl, 10 mM $(NH_4)_2SO_4$).

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. T_m 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(+) with an N-terminal FLAG tag 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 hr. The membrane was then incubated with a mouse anti-DDDDK antibody (anti-FLAG) in LI-COR Blocking Buffer for 1 hr 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 hr 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 Split-spGFP reassembly experiments were performed as previously described. N-terminal superpositive GFP tethered to

the 5-helix was cloned into pETDuet 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 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 at 37 °C for 1 hr, 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 mM 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.

ELISA Separately, wt-GLUE, GLUE-Cpep, and the C-peptide were cloned into MCS1 of pETDuet-1 with FLAG tags using restriction enzymes NcoI and NotI. The 5-helix with a C-terminal His6 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 hr 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 hr 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. 5-helix was cloned into pETDuet with an N-terminal AviTag and a C-terminal His6 tag using restriction enzymes NcoI and PacI. The completed constructs were transformed into BL21s (DE3) and purified as described previously. AviTag-5 helix was conjugated to biotin using Avidity BioMix protocols and purified BirA Protein Ligase at 1.0 mg/mL. Biotin conjugation was confirmed by Mass Spectometry. 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 hr at room temperature to block. 100 µL of Biotinylated 5-helix at a concentration of 10 µg/mL was incubated for 1 hr 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 hr at room temperature and washed 4x with 200µL wash buffer. HRP-conjugated mouse anti-FLAG antibody in LiCor blocking buffer was incubated for 1 hr at room temperature, followed by 5x 200 µL washes. Colorimetry was developed using TMB-One substrate and absorbance was measured at 655nm on a SynergyMx Microplate Reader. These ELISA data can be found in Figure S3.

Lysate Ni-NTA Pulldown Assay GLUE-Cpep was cloned into MCS1 of pETDuet-1 using restriction enzymes NcoI and NotI. The 5-helix with a C-terminal His6 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 and with 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 Spectroscopy.

2.9 Protein Used in This Work

His-GLUE

MGSSHHHHHHSQDPEYWHYVETTSSGQPLLREGEKDIFIDQSVGLYHGKSKILQRQRGR
IFLTSQRIIYIDDAKPTQNSLGLELDDLAYVNYSSGFLTRSPALILFFKDPSSSTEFVQLSFR
KSDGVLFSQATERALENILT

5-helix-His

MQLLSGIVQQQNNLLRAIEAQQHLLQLTVWGIKQLQARILAGGSGGHTTWMEWDREIN
NYTSLIHSLIEESQNQQEKNEQELLEGSSGGQLLSGIVQQQNNLLRAIEAQQHLLQLTVW
GIKQLQARILAGGSGGHTTWMEWDREINNYTSLIHSLIEESQNQQEKNEQELLEGSSGGQ
LLSGIVQQQNNLLRAIEAQQHLLQLTVWGIKQLQARILAGGHHHHHH

His-GLUE-Cpep

MGSSHHHHHHSQDPGLNDIFEAQKIEWHEGGSGGSGGTEYWHYVETTSSGQPLLREGE
KDIAIDQSVGLYHGKSKILQRQRGRIFLTSQRIIYIDDAKPTQNSLGLELDDLAYVNYSSG
FLTRSPALILFFKDPSSSTEFVQLSFRKSDGVWFSWATEIALYTILIHSLIEESQNQQEKNE
QELL

FLAG tag

MDYKDDDDK

Cpeptide

WMEWDREINNYTSLIHSLIEESQNQQEKNEQELL

CscGFP

TSGGSGKNGIKAKFKIRHNVKDGSVQLADHYQQNTPIGRGPVLLPRNHYLSTRSKLSKD
PKEKRDHMVLLEFVTAAGIKHGRDERYK

His-NscGFP

MGHHHHHHGGASKGERLFRGKVPILVELKGDVNGHKFSVRGEGKGDATRGKLTLKFIC
TTGKLPVPWPTLVTTLTYGVQCFSRYPKHMKRHDFFKSAMPKGYVQERTISFKKDGKY
KTRAEVKFEGRTLVNRIKLKGRDFKEKGNILGHKLRYNFNSHKVYITADKRGGSGSGSS

REFERENCES

1. Davis, J. M.; Tsou, L. K.; Hamilton, A. D., Synthetic non-peptide mimetics of alpha-helices. *Chem Soc Rev* **2007,** *36* (2), 326-34.

2. Ross, N. T.; Katt, W. P.; Hamilton, A. D., Synthetic mimetics of protein secondary structure domains. *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences* **2010,** *368* (1914), 989-1008.

3. Angelo, N. G.; Arora, P. S., Nonpeptidic foldamers from amino acids: synthesis and characterization of 1,3-substituted triazole oligomers. *Journal of the American Chemical Society* **2005,** *127* (49), 17134-5.

4. Kritzer, J. A., Stapled peptides: Magic bullets in nature's arsenal. *Nature chemical biology* **2010,** *6* (8), 566-7.

5. Verdine, G. L.; Hilinski, G. J., Stapled peptides for intracellular drug targets. *Methods in enzymology* **2012,** *503*, 3-33.

6. Walensky, L. D.; Bird, G. H., Hydrocarbon-Stapled Peptides: Principles, Practice, and Progress. *Journal of medicinal chemistry* **2014**.

7. Wang, D.; Chen, K.; Kulp Iii, J. L.; Arora, P. S., Evaluation of biologically relevant short alpha-helices stabilized by a main-chain hydrogen-bond surrogate. *Journal of the American Chemical Society* **2006,** *128* (28), 9248-56.

8. Goodman, C. M.; Choi, S.; Shandler, S.; DeGrado, W. F., Foldamers as versatile frameworks for the design and evolution of function. *Nature chemical biology* **2007,** *3* (5), 252-262.

9. Martinek, T. A.; Fulop, F., Peptidic foldamers: ramping up diversity. *Chem Soc Rev* **2012,** *41* (2), 687-702.

10. Hill, D. J.; Mio, M. J.; Prince, R. B.; Hughes, T. S.; Moore, J. S., A field guide to foldamers. *Chem Rev* **2001,** *101* (12), 3893-4011.

11. Gellman, S. H., Foldamers: A manifesto. *Accounts Chem Res* **1998,** *31* (4), 173-180.

12. Chapman, R. N.; Dimartino, G.; Arora, P. S., A highly stable short alpha-helix constrained by a main-chain hydrogen-bond surrogate. *Journal of the American Chemical Society* **2004,** *126* (39), 12252-3.

13. Sia, S. K.; Kim, P. S., Protein grafting of an HIV-1-inhibiting epitope. *Proceedings of the National Academy of Sciences of the United States of America* **2003,** *100* (17), 9756-61.

14. Kritzer, J. A.; Zutshi, R.; Cheah, M.; Ran, F. A.; Webman, R.; Wongjirad, T. M.; Schepartz, A., Miniature protein inhibitors of the p53-hDM2 interaction. *Chembiochem* **2006,** *7* (1), 29-31.

15. Eckert, D. M.; Kim, P. S., 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), 11187-92.

16. Eckert, D. M.; Malashkevich, V. N.; Hong, L. H.; Carr, P. A.; Kim, P. S., Inhibiting HIV-1 entry: discovery of D-peptide inhibitors that target the gp41 coiled-coil pocket. *Cell* **1999,** *99* (1), 103-15.

17. Root, M. J.; Steger, H. K., HIV-1 gp41 as a target for viral entry inhibition. *Current pharmaceutical design* **2004,** *10* (15), 1805-25.

18. Teo, H.; Gill, D. J.; Sun, J.; Perisic, O.; Veprintsev, D. B.; Vallis, Y.; Emr, S. D.; Williams, R. L., ESCRT-I core and ESCRT-II GLUE domain structures reveal role for GLUE in linking to ESCRT-I and membranes. *Cell* **2006,** *125* (1), 99-111.

19. Blakeley, B. D.; Chapman, A. M.; McNaughton, B. R., Split-superpositive GFP reassembly is a fast, efficient, and robust method for detecting protein-protein interactions in vivo. *Mol Biosyst* **2012,** *8* (8), 2036-2040.

CHAPTER THREE

Helix-Grafted Pleckstrin Homology Domains Suppress HIV-1 Infection of CD4-Positive Cells[2]

3.1 Attributions

In this work, I designed the helix grafting strategies, as well as carried out molecular cloning and protein purifications. I also performed all circular dichroism, cell lysate ELISAs and co-purification experiments. Susanne Walker, a fellow graduate student, assisted in the helix grafting design and measured the K_D of our best helix grafting proteins via ELISA. Terumasa Ikeda, our post-doctoral collaborator at the University of Minnesota, performed all live HIV-1 infectivity assays.

3.2 Introduction

Chapter II describes using a PH domain from a yeast-derived protein called GLUE as a scaffold for gp41 C-peptide helix-grafted display. Surface-exposed GLUE helix residues were mutated to match those of gp41, and the helix was extended to match the native C-peptide length using the pure gp41 sequence. 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 12 hours of exposure to human serum, supporting our initial design hypothesis that positioning the ligand within a larger stably folded structure would protect against rapid degradation.[1]

---

[2]This chapter was adapted 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.

3.3 Pleckstrin Homology Domains: A General Scaffold for Helix Grafting

Although the 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 exhibit considerably diversity in helix length, total protein size, and percentage of total residues involved in the helix.[2-9] Since 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 3.1A, Supporting Information Table S3.1**). While the canonical PH fold presents a C-terminal helix, the **E**ngulfment and Cel**l** **Mo**tility (ELMO)-1 domain has the added benefit of an additional N-terminal helix, providing the option to display partially structured ligands that are helical only at one terminus.

Our initial screen for scaffold suitability was simple expression of the wild-type proteins, since 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 3.2A, Supporting Information, Figure S3.1**). The remaining ones, including our original GLUE scaffold, were grafted with the gp41 C-peptide sequence. We tested both the N- and C-terminal ELMO sites, and were satisfied to discover that both constructs expressed admirably (**Figure 3.2B, Supporting Information, Figure S3.1**). Although other sequences also worked, ELMO seemed 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.

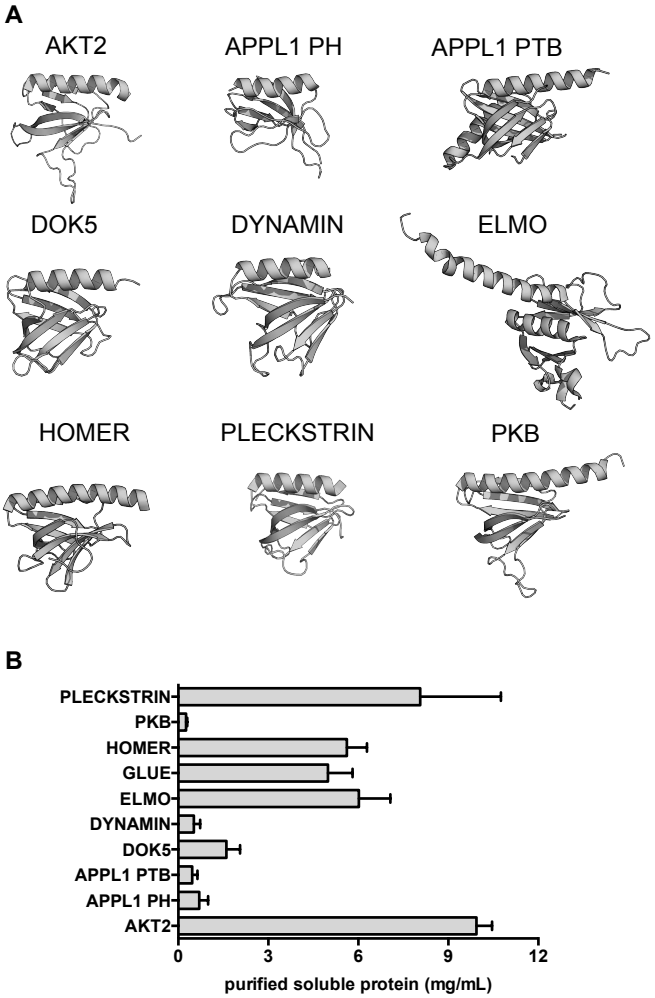
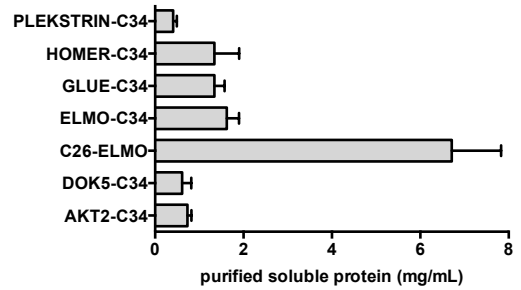


Figure 3.1 (*A*) Candidate PH domain scaffolds examined in this work. PDB: 1P6S (AKT2), 2ELA (APPL1 PH), 2ELB (APPL1 PTB), 1J0W (DOK5), 2DYN (DYNAMIN), 1I2H (HOMER), 2VSZ (ELMO),1UNP (PKB), and 2I5F (PLECKSTRIN). Not shown: 2CAY (GLUE). (*B*) Expression levels of soluble wild type scaffolds following purification by His_6/nickel NTA column. Error bars indicate standard deviation of three experiments.

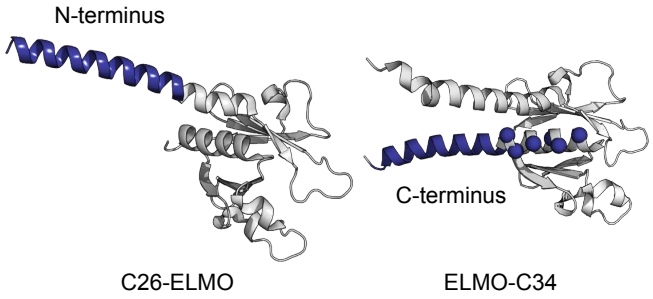
3.4 Helix Grafted ELMO

Having centered on ELMO as our most promising display vehicle, we sought to test its ability to effectively present a collection of different sequences, while at the same time probing

A



B



C

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

Figure 3.2 (*A*) Soluble expression levels (following purification by His_6/nickel NTA column purification) of selected scaffolds from Figure 2 with C-terminal C34 helix grafts, along with the N-terminal C26-ELMO graft. Error bars indicate standard deviation of three measurements. The N- and C-terminus of ELMO is labeled for clarity. (*B*) Models of N-terminal (C26-ELMO) and C-terminal (ELMO-C34) helix grafts on the ELMO scaffold. Helix grafted display strategy on the N- and C-terminal helix of ELMO (spheres indicate alpha carbon positions). (*C*) Sequence of ligand mimic region for N- and C-terminal helix-grafted ELMO (designated L-ELMO and ELMO-L, respectively).

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 the same region of gp41 as enfuvirtide (T-20), a significant portion of each sequence is unique. We thus sought to test both sequences, in addition to the longer one formed by their union (C46, **Figure 3.2B**, **Figure 3.2C**). 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, T-20, and C46. We were especially interested to probe the role of the N-terminal WWI triad on C34, whose sidechains bind into a deep (and highly conserved) hydrophobic groove on the NHR trimer surface. The

absence of this interaction in T20 (whose sequence is shifted more toward the C-peptide C-terminus) perhaps reduces barriers to acquired resistance.

The C-terminal grafts were prepared as before, with 5 solvent-exposed positions on the native ELMO helix mutated to correspond to those of the HIV sequence, and the C-terminus of the protein extended the appropriate amount using the pure gp41 sequence. In contrast to the C-terminal ELMO helix, which rests atop the beta sheets, the N-terminal one 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 3 additional residues were trimmed (since the fused helix is longer). Finally, as a hedge against even these shorter helices proving unworkable, we prepared C26-ELMO, in which 8 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 (**Supporting Information, Figure S3.2**), so much so that we used Δ21-ELMO as the starting point for the ELMO-Lig constructs.

3.5 Characterizing Helix Grafted ELMO Proteins

To establish that these various modifications had not compromised the overall protein fold, we examined each grafted protein using 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 are consistent with

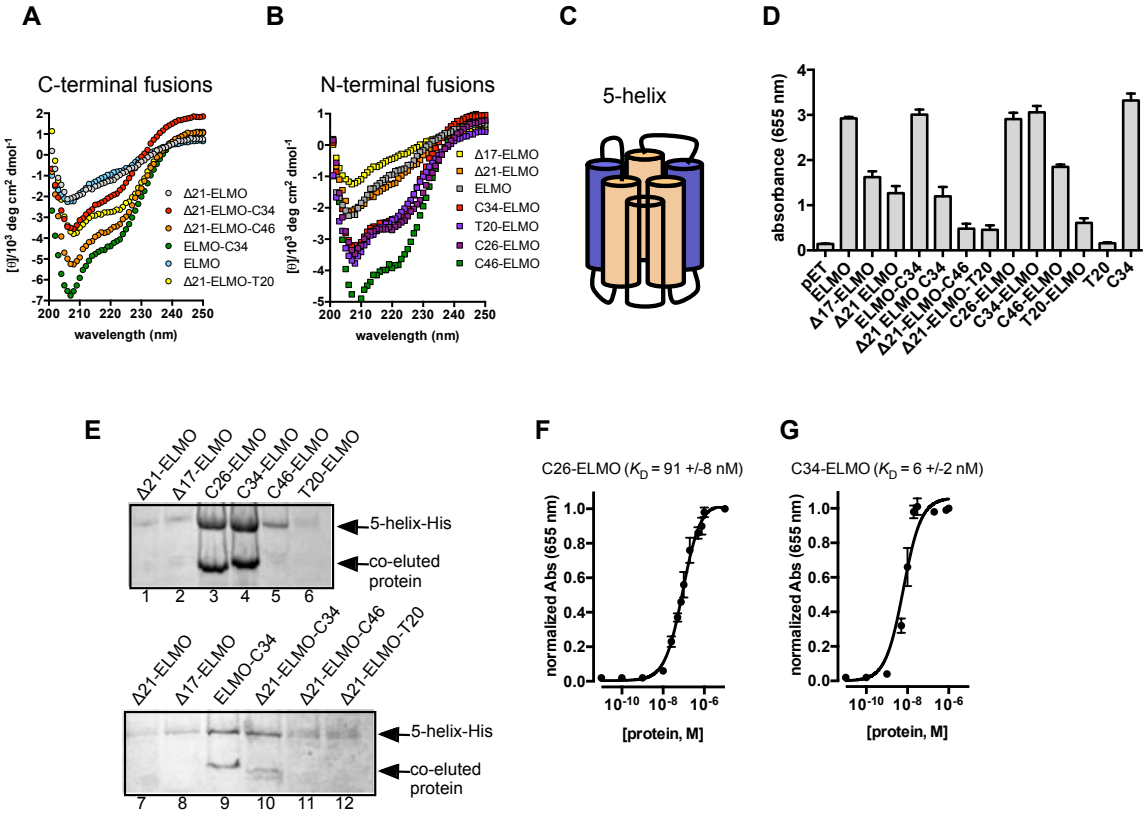


Figure 3.3 (*A*) Circular dichroism spectra of C-terminal (*A*) and N-terminal (*B*) ELMO grafts. (*C*) A cartoon representation of 5-helix, a single polypeptide that contains the NTR trimer (tan), as well as two C-peptides (purple). (*D*) ELISA data showing binding between 5-helix-His_6 and ELMO-derived grafted proteins. (*E*) Nickel-NTA co-purification of 5-helix-His and either N-terminal (top) or C-terminal (bottom) helix-grafted ELMO-derived proteins. (Dissociation constant (K_D) for complexes involving 5-helix and (*F*) C26 ELMO or (*G*) C34 ELMO.

retention of overall structure, and display expected helicity increases compared to the appropriate control (**Figure 3.3A**). Similarly, the N-terminal fusions all exhibit increased helicity with respect to each of the starting proteins (**Figure 3.3B**). The helix-grafted proteins also display similar thermal unfolding profiles, consistent with retention of macroscopic structure (**Supporting Information, Figure S3.3**). 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.

Just as Chapter II describes, initial evaluations of ligand/receptor binding employed the 5-helix construct that has been widely used for this purpose.[10, 11] The 5-helix protein links gp41's central coiled-coil N peptide trimer and two C peptide 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 (**Figure 3.3C**).

We initially performed an ELISA assay to verify binding. Somewhat surprisingly, the native N-helix on ELMO, which we originally feared might compromise expression efficiency, also has some non-specific affinity for 5-helix, at least in this assay. However, this non-specific binding can be abbrogated by truncation of the N-terminal helix. Two shortened variants, Δ17-ELMO and Δ21-ELMO (where 17 or 21 residues were removed from the N-terminal helix respectively), show substantially decreased affinity for 5-helix in ELISA (**Figure 3.3D**). The N-terminal fusions generate stronger signals, particularly those with the WWI residues: C26-ELMO and C34-ELMO are better than C46-ELMO, which in turn is better than T20-ELMO. The two shorter fusions give signals comparable to the native C34 peptide. The C-terminal fusions were weaker, with only Δ21-ELMO-C34 giving a signal even comparable to the analogous control scaffold (Δ21-ELMO).

We next examined the ability of each grafted or control protein to bind 5-helix in the complex environment of *E. coli*. Cells were transformed with a plasmid encoding both the ligand candidate and His-tagged 5-helix, lysed, and exposed to Ni-NTA agarose resin. Consistent with data points in our ELISA experiment (**Figure 3.3D**), SDS-PAGE analysis of the N-terminal fusions (**Figure 3.3E**, lanes 1-6) suggested that those bearing the WWI triad (lanes 3-5) were most effective at binding to 5-helix. 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 fusions (**Figure 3.3E**, lanes 7-12) exhibit a similar trend, albeit with an overall reduction in apparent affinity. Comparing those with the native N-helix removed (Δ 21-ELMO-Lig), the C34 fusion (lane 10) is again more effective, though the intact ELMO-C34 protein (lane 9) is perhaps even better. Both truncations (Δ17 and Δ21 ELMO do not appreciably bind 5-helix in the co-purification assay (**Figure 3.3E**)

To better understand the complex involving 5-helix and our best performing binders (C26-ELMO and C34-ELMO), we used ELISA to measure their dissociation constants (K_D). Satisfyingly, these proteins tightly bind 5-helix (K_D ~ 90 ±8 nM and 6 ±2 nM, respectively, **Figure 3.3F** and **Figure 3.3G**). In addition to tightly binding 5-helix, C26-ELMO and C34-ELMO are 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 hours (**Supporting Information, Figure S3.4**).

3.6 ELMO Suppresses HIV-1 Infection in a Live Virus Assay

Collectively, these initial experiments supported the notion that C-peptide grafted ELMO scaffolds were capable of recognizing the 5-helix model system in even 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, using a previously reported protocol.[12, 13] 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 GFP. Thus, if HIV-1 successfully infects these cells, HIV-1 Tat/TAR-dependent transcription

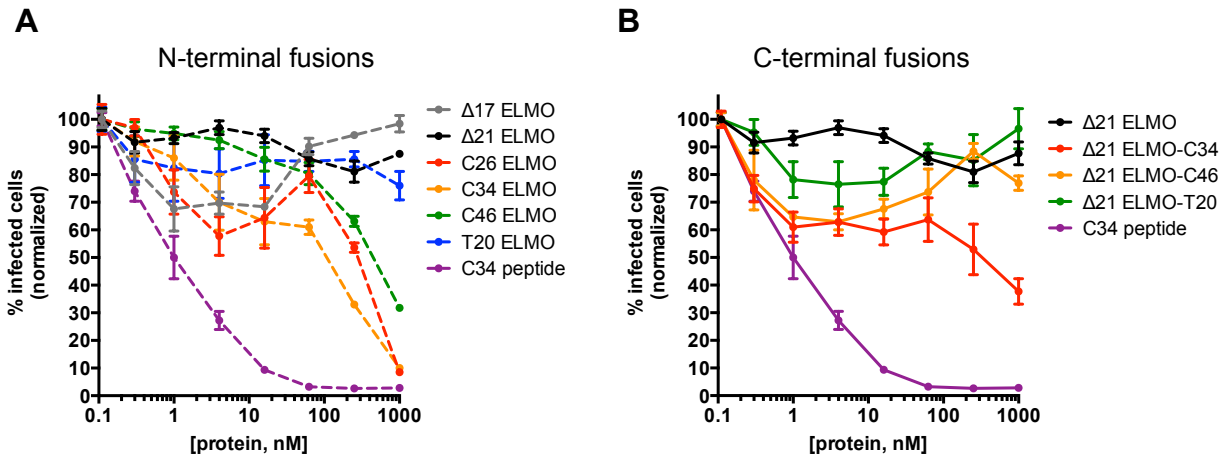


Figure 3.4. (*A*) Suppression of HIV-1 entry by ELMO-derived C-terminal fusions Δ21 ELMO (black); T20 peptide (purple); Δ21 ELMO-C34 (red); D21 ELMO-C46 (orange); D21 ELMO-T20, (green). (*B*) Suppression by N-terminal fusions Δ17 ELMO (gray); Δ21 ELMO (black); C26 ELMO (red); C34 ELMO (orange); C46 ELMO (green); T20 ELMO (blue); C34 peptide (purple).

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 fusions proved more effective by this measure as well (**Figure 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 the shorter helices, despite containing the WWI triad. The C-terminal fusions (**Figure 5B**), as before, were overall less efficient, 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 is 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 C34 peptide and its variants.

3.7 Conclusion

Taken together, these results demonstrate the viability of our helix-grafting strategy as a method for presenting gp41 C-peptide 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 may be that further reduction in scaffold size is required for true generality moving forward. Additional affinity may also be obtained through directed evolution, optimizing the choice of receptor-facing residues. Experiments along both these lines are underway. 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.

3.8 Methods

Protein Expression and Purification Genes were cloned into pET using restriction enzymes BamHI and KpnI, downstream of a His_6 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 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 (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 shown below. Purified protein concentrations were quantified using Beer's Law at an absorbance of 280 nm, following standard practice.

Resolublization of 5-Helix Inclusion Bodies 5 Helix-His_6 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-His_6 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. 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 5-12 μM in Tris buffer (20 mM Tris pH 7.4, 100 mM NaCl, 10 mM $(NH_4)_2SO_4$). 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. T_m values were obtained from minima of the first derivative of θ versus $1/T$ plots.

Serum Stability The helix-grafted C26-ELMO and C34-ELMO protein were cloned into a pETDUET upstream of a C-terminal His_6-GS-FLAG tag (DYKDDDDK). 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-FLAG antibody 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 His_6 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 protein was cloned into MCS1 of pETDuet using restriction enzymes NcoI and NotI. The 5-helix with a C-terminal His_6 tag was cloned into MCS2 of pETDuet 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. 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.

Biotinylation 5-helix-His 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). His-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-His protein at 38 μM was incubated with 6 μL of His-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 His-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 Avi-tagged 5-helix-His_6 (biotin-5-helix-His) was prepared as described above in Phosphate buffer and diluted to 10 mg/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-5-helix-His 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 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. 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.

3.9 Proteins Used in This Work

His-GLUE

MGSSHHHHHHSQDPEYWHYVETTSSGQPLLREGEKDIFIDQSVGLYHGKSKILQRQRGR
IFLTSQRIIYIDDAKPTQNSLGLELDDLAYVNYSSGFLTRSPALILFFKDPSSSTEFVQLSFR
KSDGVLFSQATERALENILT

His-GLUE-C34

MGSSHHHHHHSQDPEYWHYVETTSSGQPLLREGEKDIFIDQSVGLYHGKSKILQRQRGR
IFLTSQRIIYIDDAKPTQNSLGLELDDLAYVNYSSGFLTRSPALILFFKDPSSSTEFVQLSFR
KSDGVWFSWATEIALYTILIHSLIEESQNQQEKNEQELL

His-AKT2

MGSSHHHHHHSQDPMNEVSVIKEGWLHKRGEYIKTWRPRYFLLKSDGSFIGYKERPEA
PDQTLPPLNNFSVAECQLMKTERPRPNTFVIRCLQWTTVIERTFHVDSPDEREEWMRAIQ
MVANSLK

His-AKT2-C34

MGSSHHHHHHSQDPMNEVSVIKEGWLHKRGEYIKTWRPRYFLLKSDGSFIGYKERPEA
PDQTLPPLNNFSVAECQLMKTERPRPNTFVIRCLQWTTVIERTFHVDSPWERWEWMIAI
YTVAIHSLIEESQNQQEKNEQELL

His-APPL1 PH

MGSSHHHHHHSQDPVPDPDPTKFPVNRNLTRKAGYLNARNKTGLVSSTWDRQFYFTQ
GGNLMSQARGDVAGGLAMDIDNCSVMAVDCEDRRYCFQITSFDGKKSSILQAESKKDH
EEWICTINNISKQ

His-APPL1 PTB

MGSSHHHHHHSQDPFIVRFLGSMEVKSDDHPDVVYETMRQILAARAIHNIFRMTESHLL
VTCDCLKLIDPQTQVTRLTFPLPCVVLYATHQENKRLFGFVLRTSSGRSESNLSSVCYIFE
SNNEGEKICDSVGLAKQIALHAELDRRASEKQKEIERVK

His-DOK5

MGSSHHHHHHSQDPREQSERFNVYLMPSPNLDVHGECALQITYEYICLWDVQNPRVKLI
SWPLSALRRYGRDTTWFTFEAGRMCETGEGLFIFQTRDGEAIYQKVHSAALAIAELER

His-DOK5-C34

MGSSHHHHHHSQDPREQSERFNVYLMPSPNLDVHGECALQITYEYICLWDVQNPRVKLI
SWPLSALRRYGRDTTWFTFEAGRMCETGEGLFIFQTRDGEAIYQKVHWWAAWDIAINN
YTSLIHSLIEESQNQQEKNEQELL

His-DYNAMIN

MGSSHHHHHHSQDPILVIRKGWLTINNIGIMKGGSKEYWFVLTAENLSWYKDDEEKEK
KYMLSVDNLKLRDVEKGFMSSKHIFALFNTEQRNVYKDYRQLELACETQEEVDSWKAS
FLR

His-HOMER

MGSSHHHHHHSQDPGEQPIFSTRAHVFQIDPNTKKNWVPTSKHAVTVSYFYDSTRNVY
RIISLDGSKAIINSTITPNMTFTKTSQKFGQWADSRANTVYGLGFSSEHHLSKFAEKFQEF
KEAARLAKEKSQEK

His-HOMER-C34

MGSSHHHHHHSQDPGEQPIFSTRAHVFQIDPNTKKNWVPTSKHAVTVSYFYDSTRNVY
RIISLDGSKAIINSTITPNMTFTKTSQKFGQWADSRANTVYGLGFSSEHHLSKFAWKFWE
FKIAAYLAKIKSLEKESQNQQEKNEQELL

His-PKB

MGSSHHHHHHSQDPDVAIVKEGWLHKRGEYIKTWRPRYFLLKNDGTFIGYKERPQDVD
QREAPLNNFSVAQCQLMKTERPRPNTFIIRCLQWTTVIERTFHVETPEEREEWTTAIQTV
ADGLKKQEEEEMDFR

His-PLECKSTRIN

MGSSHHHHHHSQPGVIIKQGCLLKQGHRRKNWKVRKFILREDPAYLHYYDPAGAEDPL
GAIHLRGCVVTSVESNSNGRKSEEENLFEIITADEVHYFLQAATPKERTEWIKAIQMASR
TGKD

His-PLECKSTRIN-C34

MGSSHHHHHHSQDPGVIIKQGCLLKQGHRRKNWKVRKFILREDPAYLHYYDPAGAEDP
LGAIHLRGCVVTSVESNSNGRKSEEENLFEIITADEVHYFLQAATPWERWEWIIAIYMASI
TGLIEESQNQQEKNEQELL

His-ELMO

MGSSHHHHHHSQDPPILELKEKIQPEILELIKQQRLNRLVEGTCFRKLNARRRQDKFWYC
RLSPNHKVLHYGDLEESPQGEVPHDSLQDKLPVADIKAVVTGKDCPHMKEKGALKQN
KEVLELAFSILYDSNCQLNFIAPDKHEYCIWTDGLNALLGK

His-C26-ELMO

MGSSHHHHHHSQDPWMEWDREINNYTSLIHSLIEESQNQQKQQRLNRLVEGTCFRKLN
ARRRQDKFWYCRLSPNHKVLHYGDLEESPQGEVPHDSLQDKLPVADIKAVVTGKDCPH
MKEKGALKQNKEVLELAFSILYDSNCQLNFIAPDKHEYCIWTDGLNALLGK

His-C34-ELMO

MGSSHHHHHHSQDPWMEWDREINNYTSLIHSLIEESQNQQEKNEQELLKQQRLNRLVE
GTCFRKLNARRRQDKFWYCRLSPNHKVLHYGDLEESPQGEVPHDSLQDKLPVADIKAV
VTGKDCPHMKEKGALKQNKEVLELAFSILYDSNCQLNFIAPDKHEYCIWTDGLNALLG
K

His-C46-ELMO

MGSSHHHHHHSQDPWMEWDREINNYTSLIHSLIEESQNQQEKNEQELLELDKWASLWN
WFRLNRLVEGTCFRKLNARRRQDKFWYCRLSPNHKVLHYGDLEESPQGEVPHDSLQD
KLPVADIKAVVTGKDCPHMKEKGALKQNKEVLELAFSILYDSNCQLNFIAPDKHEYCIW
TDGLNALLGK

His-T20-ELMO

MGSSHHHHHHSQDPYTSLIHSLIEESQNQQEKNEQELLELDKWASLWNWFKQQRLNRL
VEGTCFRKLNARRRQDKFWYCRLSPNHKVLHYGDLEESPQGEVPHDSLQDKLPVADIK
AVVTGKDCPHMKEKGALKQNKEVLELAFSILYDSNCQLNFIAPDKHEYCIWTDGLNAL
LGK

His-ELMO-C34

MGSSHHHHHHSQDPPILELKEKIQPEILELIKQQRLNRLVEGTCFRKLNARRRQDKFWYC
RLSPNHKVLHYGDLEESPQGEVPHDSLQDKLPVADIKAVVTGKDCPHMKEKGALKQN
KEVLELAFSILYDSNCQLNFIAPDKWEYWIWTIGLYTLLGKSLIEESQNQQEKNEQELL

His-Δ21-ELMO

MGSSHHHHHHSQDPLNRLVEGTCFRKLNARRRQDKFWYCRLSPNHKVLHYGDLEESP
QGEVPHDSLQDKLPVADIKAVVTGKDCPHMKEKGALKQNKEVLELAFSILYDSNCQLN
FIAPDKHEYCIWTDGLNALLGK

His -Δ17-ELMO

MGSSHHHHHHSQDPKQQRLNRLVEGTCFRKLNARRRQDKFWYCRLSPNHKVLHYGDL
EESPQGEVPHDSLQDKLPVADIKAVVTGKDCPHMKEKGALKQNKEVLELAFSILYDSNC
QLNFIAPDKHEYCIWTDGLNALLGK

His-Δ21-ELMO-C34

MGSSHHHHHHSQDPLNRLVEGTCFRKLNARRRQDKFWYCRLSPNHKVLHYGDLEESP
QGEVPHDSLQDKLPVADIKAVVTGKDCPHMKEKGALKQNKEVLELAFSILYDSNCQLN
FIAPDKWEYWIWTIGLYTLLGKSLIEESQNQQEKNEQELL

His-Δ21-ELMO-C46

MGSSHHHHHHSQDPLNRLVEGTCFRKLNARRRQDKFWYCRLSPNHKVLHYGDLEESP
QGEVPHDSLQDKLPVADIKAVVTGKDCPHMKEKGALKQNKEVLELAFSILYDSNCQLN
FIAPDKWEYWIWTIGLYTLLGKSLIEESQNQQEKNEQELLELDKWASLWNWF

His- Δ21-ELMO-T20

MGSSHHHHHHSQDPLNRLVEGTCFRKLNARRRQDKFWYCRLSPNHKVLHYGDLEESP
QGEVPHDSLQDKLPVADIKAVVTGKDCPHMKEKGALKQNKEVLELAFSILYDSNCQLN
FIAPDKYEYLIWTLGLESLLGKEKNEQELLELDKWASLWNWF

5-helix-His

MQLLSGIVQQQNNLLRAIEAQQHLLQLTVWGIKQLQARILAGGSGGHTTWMEWDREIN
NYTSLIHSLIEESQNQQEKNEQELLEGSSGGQLLSGIVQQQNNLLRAIEAQQHLLQLTVW
GIKQLQARILAGGSGGHTTWMEWDREINNYTSLIHSLIEESQNQQEKNEQELLEGSSGGQ
LLSGIVQQQNNLLRAIEAQQHLLQLTVWGIKQLQARILAGGHHHHHH

FLAG tag

MDYKDDDDK

Cpeptide

WMEWDREINNYTSLIHSLIEESQNQQEKNEQELL

REFERENCES

1. Walker, S. N.; Tennyson, R. L.; 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* (2), 219-222.

2. Teo, H.; Gill, D. J.; Sun, J.; Perisic, O.; Veprintsev, D. B.; Vallis, Y.; Emr, S. D.; Williams, R. L., ESCRT-I core and ESCRT-II GLUE domain structures reveal role for GLUE in linking to ESCRT-I and membranes. *Cell* **2006,** *125* (1), 99-111.

3. Auguin, D.; Barthe, P.; Auge-Senegas, M. T.; Stern, M. H.; Noguchi, M.; Roumestand, C., Solution structure and backbone dynamics of the pleckstrin homology domain of the human protein kinase B (PKB/Akt). Interaction with inositol phosphates. *J Biomol NMR* **2004,** *28* (2), 137-55.

4. Li, J.; Mao, X.; Dong, L. Q.; Liu, F.; Tong, L., Crystal structures of the BAR-PH and PTB domains of human APPL1. *Structure* **2007,** *15* (5), 525-33.

5. Timm, D.; Salim, K.; Gout, I.; Guruprasad, L.; Waterfield, M.; Blundell, T., Crystal structure of the pleckstrin homology domain from dynamin. *Nat Struct Biol* **1994,** *1* (11), 782-8.

6. Komander, D.; Patel, M.; Laurin, M.; Fradet, N.; Pelletier, A.; Barford, D.; Cote, J. F., An alpha-helical extension of the ELMO1 pleckstrin homology domain mediates direct interaction to DOCK180 and is critical in Rac signaling. *Mol Biol Cell* **2008,** *19* (11), 4837-51.

7. Irie, K.; Nakatsu, T.; Mitsuoka, K.; Miyazawa, A.; Sobue, K.; Hiroaki, Y.; Doi, T.; Fujiyoshi, Y.; Kato, H., Crystal structure of the Homer 1 family conserved region reveals the interaction between the EVH1 domain and own proline-rich motif. *J Mol Biol* **2002,** *318* (4), 1117-26.

8. Jackson, S. G.; Zhang, Y.; Haslam, R. J.; Junop, M. S., Structural analysis of the carboxy terminal PH domain of pleckstrin bound to D-myo-inositol 1,2,3,5,6-pentakisphosphate. *BMC Struct Biol* **2007,** *7*, 80.

9. Milburn, C. C.; Deak, M.; Kelly, S. M.; Price, N. C.; Alessi, D. R.; Van Aalten, D. M., Binding of phosphatidylinositol 3,4,5-trisphosphate to the pleckstrin homology domain of protein kinase B induces a conformational change. *Biochem J* **2003,** *375* (Pt 3), 531-8.

10. Root, M. J.; Kay, M. S.; Kim, P. S., Protein design of an HIV-1 entry inhibitor. *Science* **2001,** *291* (5505), 884-8.

11. Horne, W. S.; Johnson, L. M.; Ketas, T. J.; Klasse, P. J.; Lu, M.; Moore, J. P.; Gellman, S. H., Structural and biological mimicry of protein surface recognition by alpha/beta-peptide foldamers. *Proc Natl Acad Sci U S A* **2009,** *106* (35), 14751-6.

12. Gervaix, A.; West, D.; Leoni, L. M.; Richman, D. D.; Wong-Staal, F.; Corbeil, J., A new reporter cell line to monitor HIV infection and drug susceptibility in vitro. *Proc Natl Acad Sci U S A* **1997,** *94* (9), 4653-8.

13. Hache, G.; Shindo, K.; Albin, J. S.; Harris, R. S., Evolution of HIV-1 isolates that use a novel Vif-independent mechanism to resist restriction by human APOBEC3G. *Curr Biol* **2008,** *18* (11), 819-24.

CHAPTER FOUR

Evaluation of Sequence Variability in HIV-1 gp41 C-peptide Helix-Grafted Proteins[3]

4.1 Attributions

In this work, I assisted in molecular cloning and protein purification. I also carried out the design, execution and analysis of all circular dichroism experiments. Susanne Walker, a fellow graduate student constructed the design and screening of the protein library. Terumasa Ikeda, our post-doctoral collaborator at the University of Minnesota, performed all live HIV-1 infectivity assays.

4.1 Introduction

We have used helix-grafted display to create proteins capable of suppressing HIV entry.[1, 2] 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 as described in Chapter I. Helix-grafted proteins presenting a C-peptide surrogate bind pre-fusogenic gp41 N-peptide trimer, leading to suppression of gp41 fusogenic assembly. Our second-generation helix grafted protein, 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 4.1A**). 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.[2] In

[3]This chapter was adapted 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. *Bioorg Med Chem* **2018,** *26* (6), 1220-1224.

contrast to its peptide counterpart (the FDA approved peptide drug Fuzeon™), Cpep-ELMO is stable in human serum.

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.

4.2 Yeast Display Evolution of a Cpep-ELMO Helix-Grafted Display Protein Library

As stated above, we 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.[2] We next sought to evaluate sequence variability in the grafted helix. Five positions on Cpep-ELMO (**Figure 4.1A**) 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 4.1A)** we mutated are buried in a deep (and conserved) hydrophobic pocket; their deletion has been reported to significantly reduce

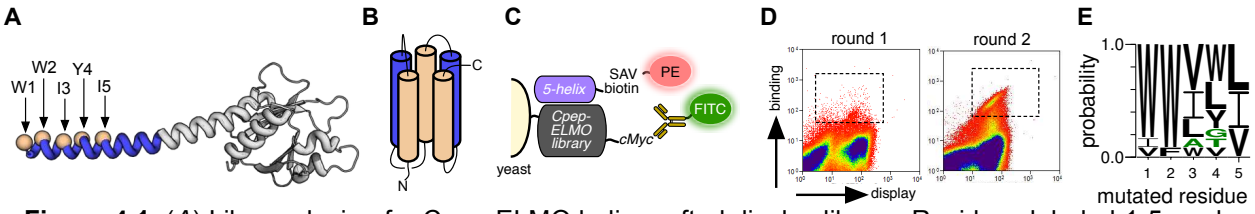


Figure 4.1. (*A*) Library design for Cpep-ELMO helix-grafted display library. Residues labeled 1-5, and highlighted as light brown spheres, were mutated to all possible 20 proteinogenic amino acids, generating a ~3.2 million-member protein library. (*B*) Cartoon depiction of gp41 5-helix, a single polypeptide that contains three N-peptides (brown) and two C-peptides (blue), which presents a single available C-peptide binding site. (*C*) Flow cytometry enrichment scheme to identify novel Cpep-ELMO helix-grafted proteins that bind to gp41 5-helix. (*D*) Flow cytometry data from yeast display library screening (round 1: left; round 2: right). Yeast selected in round 1 and round 2, representing 0.4% and 1% of the total population, respectively, are boxed. (*E*) Sequence logo generated from 20 Cpep-ELMO derived mutants enriched after two rounds of yeast display screening.

affinity[3]. 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 an C-terminal *cMyc* tag. Following display, yeast was concomitantly incubated with a FITC-labelled anti-*cMyc* antibody and biotinylated gp41 5-helix. 5-helix is a previously reported protein that covalently tethers five of the six subunits within the gp41 trimer-of-hairpins (three N-peptides and two C-peptides) using simple GlyGlySer connectors.[4, 5] When folded, it features the gp41 coiled coil with two of its binding sides already occupied, and just a single exposed interface (**Figure 4.1B**). 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 added. 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 4.1C**). 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 4.1D**).

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 (**Figure 4.1A**) 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 4.1E**). 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.

4.3 Evolved Cpep-ELMO Mutants Retain Affinity for HIV-1 gp41 5-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 Enzyme-Linked Immunosorbent Assay (ELISA).

Biotinylated gp41 5-helix 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 binding as then analyzed by measuring luminescence, following incubation with necessary ELISA reagents. Of the initial eleven hits, five proteins (Cpep1, 3, 4, 5, 8-ELMO) compared favorably to Cpep-ELMO and warranted further investigation (**Figure**

4.2A). Amino acid composition of Cpep-ELMO, and our evolved proteins at mutated positions is shown in **Figure 4.2B**.

4.4 Evolved Cpep-ELMO Proteins are Structured, Stable, and Express Well in *E. coli*

Given the findings from our yeast display screening and ELISA experiments, whether or not solvent exposed residues on the gp41 C-peptide helix are amenable to mutation is no longer in question. What remained in question, however, was what affect these mutations had on the larger protein structure. We began to answer this question 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 protein stability and expression[6]. We expressed five of the tightest binding Cpep-ELMO mutants in *E. coli* as N-terminal His_6 fusions, and purified them by nickel-NTA chromatography. Levels of purified eluted proteins were then measured as mg/L with spectroscopy (nanodrop analysis). As shown in **Figure 4.2C**, all but one of our evolved proteins express at higher levels, compared to Cpep-ELMO. While Cpep-ELMO expresses modestly in *E. coli* (~2.5 mg/mL), our best performing mutants, Cpep3-ELMO and Cpep4-ELMO, express more than twice as well.

In addition to expressing better in *E. coli*, evolved mutants retain structural features associated with our starting protein (**Figure 4.2D**) as assed 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

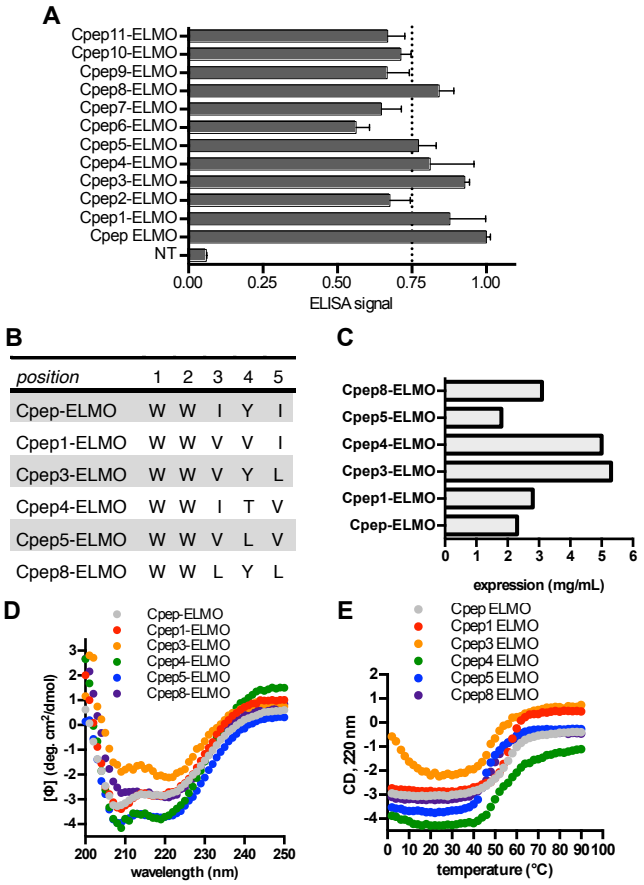


Figure 4.2. (*A*) ELISA data showing binding between immobilized gp41 5-helix and Cpep-ELMO, or mutants described in this work. Error bars represent the standard error of the mean from three replicate experiments. (*B*) Sequence of Cpep-ELMO, and mutants selected for by yeast display in this work (Cpep1-ELMO – Cpep11-ELMO). (*C*) Expression levels, in *E. coli*, for our starting helix-grafted display protein (Cpep-ELMO) and mutants generated from our yeast display screen (Cpep1-11-ELMO). (*D*) Circular dichroism spectra for Cpep-ELISA and mutants described in this work. (*E*) Melting data for Cpep-ELMO and mutants described in this work

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 4.2E**). 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.

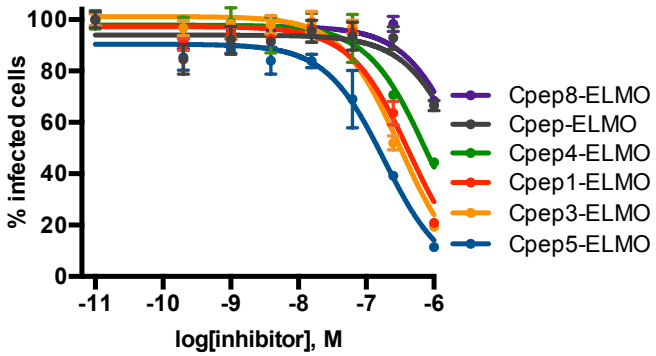


Figure 4.3. Suppression of HIV-1 entry in CD4-positive cells in a live virus assay by Cpep-ELMO, and evolved variants thereof. Error bars represent the standard error of the mean from three replicate experiments

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[2, 7]. 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 4.3**, all but one (Cpep8-ELMO) of the evolved helix-grafted display proteins inhibit HIV-1 entry better than the original helix-grafted display protein, Cpep-ELMO. A number of our mutated proteins have dramatically improved Inhibitory Constant-50 (IC_{50}) values for inhibition 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.

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.

Recently we reported 'helix-grafted display', a potentially general solution for displaying a folded therapeutically-relevant helix on a protein scaffold. In recently reported work, 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. To date, our research has focused on the development of new proteins that display HIV-1 gp41 C-peptide, an established therapeutic reagent. Previously, 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 4.1a**) 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 4.1a**) 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.

4.7 Methods

Protein Expression and Purification Genes were cloned into pET using restriction enzymes BamHI and KpnI, downstream of a His_6 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 mM 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. Protein expression was calculated as a measure of eluted protein yield in milligrams per liter of induced *E. coli* culture.

Resolublization of 5-helix Inclusion Bodies 5-helix-His_6 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-His 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 using 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.

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. T_m values were obtained from minima of the first derivative of θ versus $1/T$ plots.

Biotinylation 5-helix-His 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. His-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-His_6 protein at 38 μM was incubated with 6 mL of His_6-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 His-GS-FLAG tag (DYKDDDDK) using restriction enzymes NheI and BamHI. Proteins were purified as described above in PBS buffer. The biotinylated Avitagged^{TM} 5-helix-His_6 (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 mL 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_2HPO4, 8.6 g/L $NaH_2PO_4 \bullet H_2O$, 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 $1 \times 10^7 - 1.63 \times 10^7$.

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 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 SAPE 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 (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 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 Proteins Used in This Work

5-helix-His

MQLLSGIVQQQNNLLRAIEAQQHLLQLTVWGIKQLQARILAGGSGGHTTWMEWDREIN
NYTSLIHSLIEESQNQQEKNEQELLEGSSGGQLLSGIVQQQNNLLRAIEAQQHLLQLTVW
GIKQLQARILAGGSGGHTTWMEWDREINNYTSLIHSLIEESQNQQEKNEQELLEGSSGGQ
LLSGIVQQQNNLLRAIEAQQHLLQLTVWGIKQLQARILAGGHHHHHH*

His-ELMO

MGSSHHHHHHSQDPKQQRLNRLVEGTCFRKLNARRRQDKFWYCRLSPNHKVLHYGDL
EESPQGEVPHDSLQDKLPVADIKAVVTGKDCPHMKEKGALKQNKEVLELAFSILYDSNC
QLNFIAPDKHEYCIWTDGLNALLGK*

His-Cpep-ELMO

MGSSHHHHHHSQDPWMEWDREINNYTSLIHSLIEESQNQQKQQRLNRLVEGTCFRKLN
ARRRQDKFWYCRLSPNHKVLHYGDLEESPQGEVPHDSLQDKLPVADIKAVVTGKDCPH
MKEKGALKQNKEVLELAFSILYDSNCQLNFIAPDKHEYCIWTDGLNALLGK*

His-Cpep1-ELMO

MGSSHHHHHHSQDPWMEWDREVNNVTSLIHSLIEESQNQQKQQRLNRLVEGTCFRKLN
ARRRQDKFWYCRLSPNHKVLHYGDLEESPQGEVPHDSLQDKLPVADIKAVVTGKDCPH
MKEKGALKQNKEVLELAFSILYDSNCQLNFIAPDKHEYCIWTDGLNALLGK*

His-Cpep2-ELMO

MGSSHHHHHHSQDPWMEWDRELNNWTSLIHSLIEESQNQQKQQRLNRLVEGTCFRKL
NARRRQDKFWYCRLSPNHKVLHYGDLEESPQGEVPHDSLQDKLPVADIKAVVTGKDCP
HMKEKGALKQNKEVLELAFSILYDSNCQLNFIAPDKHEYCIWTDGLNALLGK*

His-Cpep3-ELMO

MGSSHHHHHHSQDPWMEWDREVNNYTSLLHSLIEESQNQQKQQRLNRLVEGTCFRKL
NARRRQDKFWYCRLSPNHKVLHYGDLEESPQGEVPHDSLQDKLPVADIKAVVTGKDCP
HMKEKGALKQNKEVLELAFSILYDSNCQLNFIAPDKHEYCIWTDGLNALLGK*

His-Cpep4-ELMO

MGSSHHHHHHSQDPWMEWDREINNTTSLVHSLIEESQNQQKQQRLNRLVEGTCFRKLN
ARRRQDKFWYCRLSPNHKVLHYGDLEESPQGEVPHDSLQDKLPVADIKAVVTGKDCPH
MKEKGALKQNKEVLELAFSILYDSNCQLNFIAPDKHEYCIWTDGLNALLGK*

His-Cpep5-ELMO

MGSSHHHHHHSQDPWMEWDREVNNLTSLVHSLIEESQNQQKQQRLNRLVEGTCFRKL
NARRRQDKFWYCRLSPNHKVLHYGDLEESPQGEVPHDSLQDKLPVADIKAVVTGKDCP
HMKEKGALKQNKEVLELAFSILYDSNCQLNFIAPDKHEYCIWTDGLNALLGK*

His-Cpep6-ELMO

MGSSHHHHHHSQDPIMEWDREVNNWTSLLHSLIEESQNQQKQQRLNRLVEGTCFRKLN
ARRRQDKFWYCRLSPNHKVLHYGDLEESPQGEVPHDSLQDKLPVADIKAVVTGKDCPH
MKEKGALKQNKEVLELAFSILYDSNCQLNFIAPDKHEYCIWTDGLNALLGK*

His-Cpep7-ELMO

MGSSHHHHHHSQDPWMEFDREANNLTSLIHSLIEESQNQQKQQRLNRLVEGTCFRKLN
ARRRQDKFWYCRLSPNHKVLHYGDLEESPQGEVPHDSLQDKLPVADIKAVVTGKDCPH
MKEKGALKQNKEVLELAFSILYDSNCQLNFIAPDKHEYCIWTDGLNALLGK*

His-Cpep8-ELMO

MGSSHHHHHHSQDPWMEWDRELNNYTSLLHSLIEESQNQQKQQRLNRLVEGTCFRKL
NARRRQDKFWYCRLSPNHKVLHYGDLEESPQGEVPHDSLQDKLPVADIKAVVTGKDCP
HMKEKGALKQNKEVLELAFSILYDSNCQLNFIAPDKHEYCIWTDGLNALLGK*

His-Cpep9-ELMO

MGSSHHHHHHSQDPWMEWDREINNWTSLLHSLIEESQNQQKQQRLNRLVEGTCFRKL
NARRRQDKFWYCRLSPNHKVLHYGDLEESPQGEVPHDSLQDKLPVADIKAVVTGKDCP
HMKEKGALKQNKEVLELAFSILYDSNCQLNFIAPDKHEYCIWTDGLNALLGK*

His-Cpep10-ELMO

MGSSHHHHHHSQDPWMEWDREINNYTSLIHSLIEESQNQQKQQRLNRLVEGTCFRKLN
ARRRQDKFWYCRLSPNHKVLHYGDLEESPQGEVPHDSLQDKLPVADIKAVVTGKDCPH
MKEKGALKQNKEVLELAFSILYDSNCQLNFIAPDKHEYCIWTDGLNALLGK*

His-Cpep11-ELMO

MGSSHHHHHHSQDPVMEWDREVNNWTSLIHSLIEESQNQQKQQRLNRLVEGTCFRKLN
ARRRQDKFWYCRLSPNHKVLHYGDLEESPQGEVPHDSLQDKLPVADIKAVVTGKDCPH
MKEKGALKQNKEVLELAFSILYDSNCQLNFIAPDKHEYCIWTDGLNALLGK*

His-BirA

MKDNTVPLKLIALLANGEFHSGEQLGETLGMSRAAINKHIQTLRDWGVDVFTVPGKGY
SLPEPIQLLNAKQILGQLDGGSVAVLPVIDSTNQYLLDRIGELKSGDACIAEYQQAGRGR
RGRKWFSPFGANLYLSMFWRLEQGPAAAIGLSLVIGIVMAEVLRKLGADKVRVKWPND
LYLQDRKLAGILVELTGKTGDAAQIVIGAGINMAMRRVEESVVNQGWITLQEAGINLDR
NTLAAMLIRELRAALELFEQEGLAPYLSRWEKLDNFINRPVKLIIGDKEIFGISRGIDKQG
ALLLEQDGIIKPWMGGEISLRSAEKGSHHHHHH*

REFERENCES

1. Walker, S. N.; Tennyson, R. L.; 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* (2), 219-22.

2. 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* (20), 1945-1950.

3. Chan, D. C.; Chutkowski, C. T.; Kim, P. S., Evidence that a prominent cavity in the coiled coil of HIV type 1 gp41 is an attractive drug target. *Proc Natl Acad Sci U S A* **1998,** *95* (26), 15613-7.

4. Root, M. J.; Kay, M. S.; Kim, P. S., Protein design of an HIV-1 entry inhibitor. *Science* **2001,** *291* (5505), 884-8.

5. Horne, W. S.; Johnson, L. M.; Ketas, T. J.; Klasse, P. J.; Lu, M.; Moore, J. P.; Gellman, S. H., Structural and biological mimicry of protein surface recognition by alpha/beta-peptide foldamers. *Proc Natl Acad Sci U S A* **2009,** *106* (35), 14751-6.

6. Boder, E. T.; Wittrup, K. D., Yeast surface display for directed evolution of protein expression, affinity, and stability. *Methods Enzymol* **2000,** *328*, 430-44.

7. Hache, G.; Shindo, K.; Albin, J. S.; Harris, R. S., Evolution of HIV-1 isolates that use a novel Vif-independent mechanism to resist restriction by human APOBEC3G. *Curr Biol* **2008,** *18* (11), 819-24.

CHAPTER FIVE

Evolved Proteins Inhibit Entry of Enfuvirtide-Resistant HIV-1[4]

5.1 Attributions

In this work, I designed the helix grafting strategies, as well as carried out molecular cloning and protein purifications. I also performed all circular dichroism, cell lysate ELISAs and co-purification experiments. Susanne Walker, a fellow graduate student, constructed the design and screening of the protein library. Terumasa Ikeda, our post-doctoral collaborator at the University of Minnesota, performed all live HIV-1 infectivity assays.

5.2 Introduction

Previous studies suggested a relationship between the size of the helix-grafted display scaffold and HIV-1 entry inhibition: smaller scaffolds are more effective.[1] 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).[2]

[4] This chapter was adapted from:
Tennyson, R. L.; Walker, S. N.; Ikeda, T.; Harris, R. S.; McNaughton, B. R., Evolved Proteins Inhibit Entry of Enfuviritde-Resistant HIV-1. *ACS Infect. Dis* **Just Accepted**.

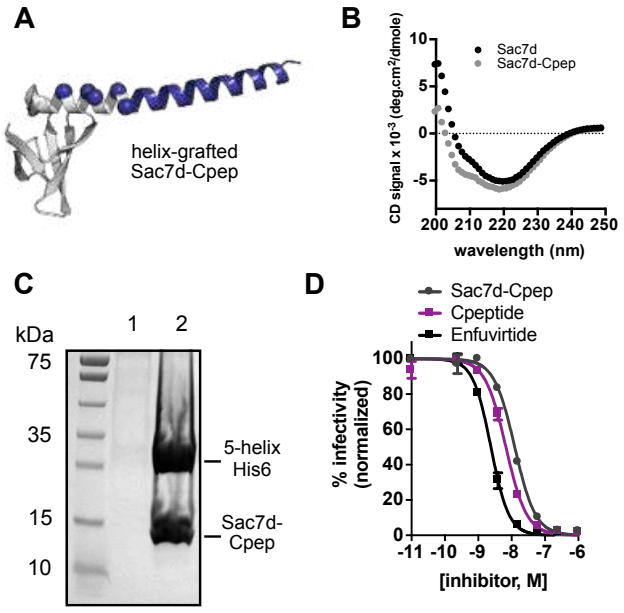


Figure 5.1 (*A*) Helix-grafted Sac7d-Cpep. Blue represents grafted C-peptide residues. Gray represents Sac7d-residues. (*B*) Circular dichroism data for Sac7d and Sac7d-Cpep. (*C*) Sac7d-Cpep binds 5-helix in *E. coli*. Lane 1: *E. coli* cell lysate containing Sac7d and 5-helix-His6. Sac7d does not co-purify 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*) Sac7d-Cpep potently inhibits HIV entry in a live virus assay (IC_{50} 1.9-12.4 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.

Sac7d was engineered, using helix-grafted display, to function as a C-peptide mimic, and the resulting protein (Sac7d-Cpep) was analyzed for structural and functional fidelity. Sac7d-Cpep (**Figure 5.1A**) expresses as a soluble protein in *E. coli*, and circular dichroism experiments show that it maintains structural features found in the wild-type protein (**Figures 5.1B and S5.1A**). When 5-helix-His6 is expressed in *E. coli* in the absence of a binding partner, it expressed in the form of an inclusion body, that must be refolded. This is likely due to aggregation caused by the solvent-exposed hydrophobic C-peptide binding site. However, when 5-helix-His6 is appropriately bound by another protein (lacking a His_6 tag), effectively filling the hydrophobic C-peptide binding site, both proteins can be co-purified from *E. coli* lysate with nickel-NTA agarose beads. When 5-helix-His6 and Sac7d are co-expressed in *E. coli*, we

observe no purified protein following incubation with nickel-NTA agarose beads and elution with imidazole (**Figure 5.1C**, lane 1), indicating that Sac7d does not appreciably bind 5-helix-His6. However, when 5-helix-His is expressed concomitantly with Sac7d-Cpep, both proteins readily co-purify following incubation with nickel-NTA agarose beads and elution with imidazole (**Figure 5.1C**, 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 ELISA, which indicates binding between Sac7d-Cpep and immobilized 5-helix, but no appreciable affinity between Sac7d and 5-helix (**Figure S5.1B**).

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.[3, 4] 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).[5] 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 (data not shown). In contrast, Sac7d-Cpep potently inhibits HIV entry (IC_{50} 1.9-12.4 nM, **Figure 5.1D**; **Table 5.1**, entry 1), and compares similarly to Enfuvirtide (IC_{50} 0.7-3.5 nM over 6 different experiments over the course of this study, each performed in triplicate, **Figure 5.1D**; **Table 5.1**, entry 2) and gp41 C-peptide (IC_{50} 1.0-6.7 nM, **Figure 5.1D**; **Table 5.1**, entry 3). 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 sequences (**Figure S1C** and **Table S5.2**). 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 tryptophan residues necessary for 5-helix recognition, this scaffold cannot be evolved for improved 5-helix recognition. Given these results, we focused on Sac7d-Cpep as a starting point to generate peptides with improved therapeutic properties including affinity, solubility, and durability.

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 diverse set 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 an ensemble consisting of mixed proteins, 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 Enfuvirtide resistance.

The first two C-peptide grafted residues on Sac7d-Cpep are both tryptophans (WW, white spheres, **Figure 5.2A**). In the context of helix-grafted proteins, we have demonstrated these residues are necessary for recognition of 5-helix. [3] Thus, we avoided mutating these tryptophans in our library. In the first-generation of helix optimization, we mutated five residues (grafted and native) on the solvent-exposed face of the helix (EIYTI, light blue spheres, **Figure**

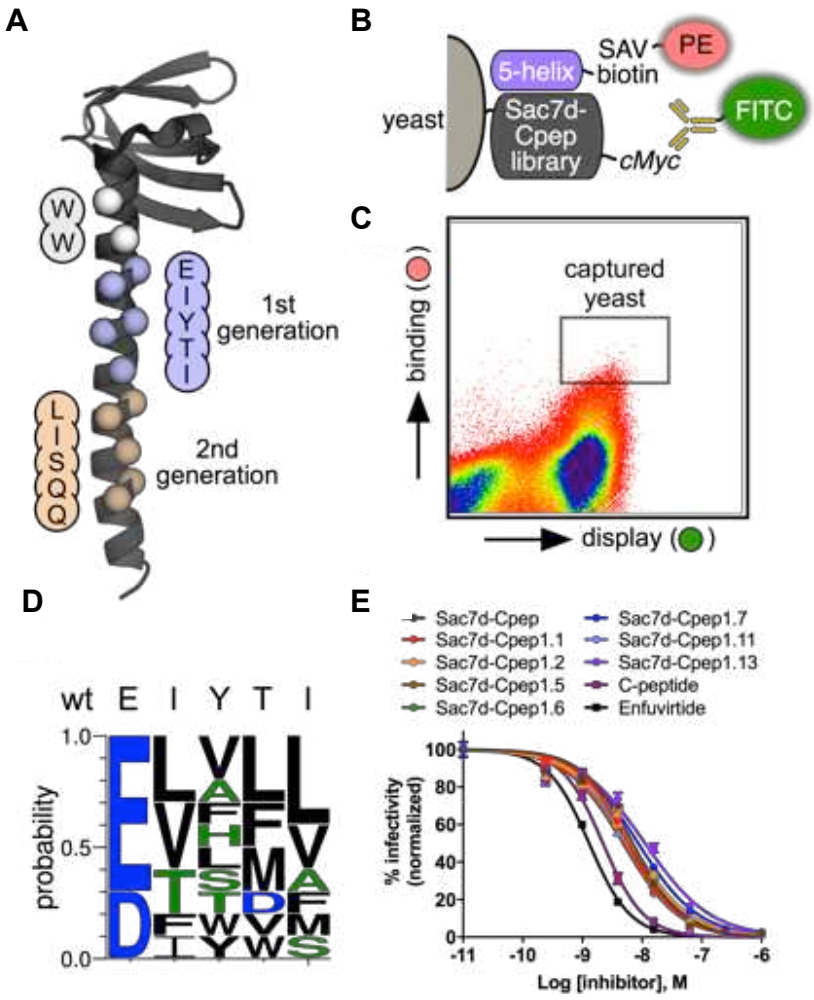


Figure 5.2 (*A*) Helix optimization strategy. Light blue residues were optimized in generation 1; tan residues were optimized in generation 2. (*B*) Yeast display optimization assay. (*C*) Representative yeast display data from generation 1 screening. (*D*) Sequence logo for residues optimized during generation 1 of yeast display screening. (*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.

5.2A) to all 20 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. As in our prior work, yeast were concomitantly incubated with a FITC-labeled anti-Myc antibody and biotinylated gp41 5-helix. Following a brief washing step, streptavidin-phycoerythrin (SAV-PE) conjugate was added. Thus, yeast featuring biotin near their cell

surface, by virtue of a surface protein/5-helix interaction, will be bound by SAV-PE (**Figure 5.2B**). 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 S5.2A and Table S5.1**). A representative flow cytometry experiment (1 nM 5-helix incubation with 10 nM Cpep-ELMO competitor) is shown in Figure 3C.

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 3D. For the first randomized residue (glutamic acid, E, **Figure 5.2A**), we observe a preference for the wild-type amino acid. Aspartic acid was also observed in selected proteins, suggesting that a carboxylic acid containing side chain is favored at this position. However, significant heterogeneity was observed 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 ELISA (**Figure S5.2B**) and seven proteins were identified as competitive with Sac7d-Cpep (greater than 90% ELISA signal was generated, 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 5.2E**, **Table 5.1**, entry 4-10). To differentiate between these proteins, we compared their expression yield in *E. coli* (**Figure S5.2C**). We observed a 18-fold increase in expression of Sac7d-Cpep1.1,

compared to Sac7d-Cpep. Residues optimized in Sac7d-Cpep1.1 are EAWLL, compared to wild-type residues EIYTI; this protein was used as a starting point for second-generation evolution.

Table 5.1 Sequence of HIV entry inhibitors, and associated half maximal inhibitory constant (IC_{50}) values

entry	protein			generation 1					generation 2					HIV IIIB (IC50, nM)	V38A/N42D (IC50, nM)
1	Sac7d-Cpep	W	W	E	I	Y	T	I	L	I	S	Q	Q	1.9-12.4 +/- 0.2-1.3	5.4 +/- 0.4
2	Cpeptide	W	W	E	I	Y	T	I	L	I	S	Q	Q	1.0-6.7 +/- 0.1-0.2	72.9 +/- 12.4
3	Enfuvirtide					Y	T	I	L	I	S	Q	Q	0.7-3.5 +/- 0.1-0.2	401.3 +/- 47.4
4	Sac7d-Cpep1.1	W	W	E	A	W	L	L	L	I	S	Q	Q	2.6-6.2 +/- 0.3-0.6	2.1 +/- 0.2
5	Sac7d-Cpep1.2	W	W	E	Y	L	I	L	L	I	S	Q	Q	7.7 +/- 1.4	-
6	Sac7d-Cpep1.5	W	W	E	I	A	L	V	L	I	S	Q	Q	5.4 +/- 0.9	-
7	Sac7d-Cpep1.6	W	W	E	T	V	M	L	L	I	S	Q	Q	7.0 +/- 1.2	-
8	Sac7d-Cpep1.7	W	W	E	L	S	W	L	L	I	S	Q	Q	9.4 +/- 1.4	-
9	Sac7d-Cpep1.11	W	W	D	V	L	L	M	L	I	S	Q	Q	5.8 +/- 1.0	-
10	Sac7d-Cpep1.13	W	W	E	L	Y	F	S	L	I	S	Q	Q	15.6 +/- 4.0	-
11	Sac7d-Cpep2.5	W	W	E	A	W	L	L	I	V	L	T	L	29.0 +/- 3.5	30.5 +/- 8.7
12	Sac7d-Cpep2.8	W	W	E	A	W	L	L	L	F	L	A	Q	4.1 +/- 0.8	15.9 +/- 2.3
13	Sac7d-Cpep2.10	W	W	E	A	W	L	L	L	L	M	Q	Q	5.1 +/- 1.5	4.3 +/- 0.6
14	Sac7d-Cpep2.13	W	W	E	A	W	L	L	L	I	G	M	S	13.3 +/- 3.4	55.0 +/- 7.0
15	Sac7d-Cpep2.14	W	W	E	A	W	L	L	I	L	I	E	I	20.7 +/- 7.9	83.7 +/- 16.7
16	Sac7d-Cpep2.19	W	W	E	A	W	L	L	L	L	A	R	H	5.7 +/- 1.1	11.7 +/- 1.8
17	Sac7d-Cpep2.20	W	W	E	A	W	L	L	F	L	A	S	Q	4.6 +/- 0.6	28.2 +/- 4.1
18	mFc-Sac7d-Cpep nl	W	W	E	I	Y	T	I	L	I	S	Q	Q	282.9 +/- 17.8	-
19	mCH3-Sac7d-Cpep nl	W	W	E	I	Y	T	I	L	I	S	Q	Q	18.0 +/- 4.8	-
20	SAbp-Sac7d-Cpep nl	W	W	E	I	Y	T	I	L	I	S	Q	Q	3.9 +/- 0.8	-

5.5 Second-Generation Helix Evolution Delivers a Library of Potent Entry Inhibitors

Solvent-exposed residues on the C-terminal region of the grafted helix of Sac7d-Cpep1.1 (LISQQ, tan spheres, **Figure 5.2A**) 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 S5.3A** and **Table S5.1**).

Following the final enrichment round, we sequenced 50 plasmids that encode for 20 unique protein sequences. We observed significant sequence variability (**Figure 5.3A**), suggesting that multiple solutions to potent 5-helix recognition exist within this region of the C-

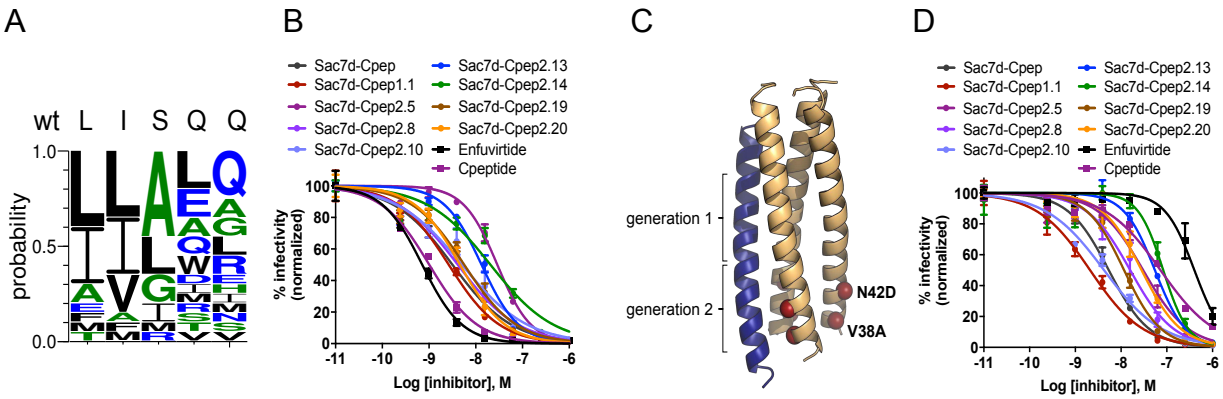


Figure 5.3. (*A*) Sequence logo for residues optimized during generation 2 of yeast display screening (*B*) Inhibition of HIV entry of Sac7d-Cpep, C-peptide, Enfuvirtide, and generation 2 optimized proteins. Error bars represent the standard error of the mean (SEM) for three separate experiments. If error bars are not visible, error is smaller than the data point. (*C*) Position of V38A and N42D mutations within the N-heptad repeat (brown) of gp41. Mutations are highlighted in red. C-peptide is highlighted in blue; regions evolved generationally are shown. (*D*) Inhibition of 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.

peptide helix. Affinity for 5-helix recognition was again assessed by ELISA (**Figure S5.3B**) and lowest affinity binders were triaged. Evolved proteins with low expression in *E. coli* were also eliminated (**Figure S5.3C**). Entry inhibition was measured for seven second-generation proteins (**Figure 5.3B**). Despite substantial differences in their sequences, all evaluated proteins potently inhibit HIV entry (IC_{50} 4.1-29.0 nM, Table 1, entry 11-17). Moreover, despite dramatic resurfacing of the solvent-exposed face of C-peptide, proteins evolved in second-generation helix optimization expressed 2-9-fold more efficiently in *E. coli*, compared to Sac7d-Cpep (**Figure S5.3C**).

5.6 Evolved Proteins Bind Enfuvirtide-Resistant 5-helix and Inhibit Entry of Enfuvirtide-Resistant HIV

Given the sequence diversity of the evolved helix, we reasoned that at least some of the evolved proteins might recognize 5-helix (and thus gp41) in slightly different ways, and/or 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 the entry of a mutant form of gp41 that endows resistance to Enfuvirtide.

Enfuvirtide is given to patients as a last resort therapy due to, at least in part, rapid generation of resistance due to mutation of gp41. In patient-derived studies, a frequent gp41 mutation within the N-terminal heptad repeat (NHR, **Figure 5.3C**, brown) region that endows Enfuvirtide resistance is V38A (**Figure S5.4A)**. Additionally, double mutations within the same region (V38A/N42D; **Figure 5.3C**) have been found to confer cross-resistance to Enfuvirtide and C-peptide.[6] To determine if our evolved proteins are functional against these clinically relevant forms of HIV, we generated Enfuvirtide-resistant HIV IIIB virions with the V38A single substitution or V38A/N42D double mutation. As in the live virus assay with the wild-type 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.

All eight of the evolved proteins we tested potently inhibited entry of the Enfuvirtide-resistant double mutant strain (IC_{50} 4.3-83.7 nM), and compared favorably to C-peptide (IC_{50} 72.9 nM; **Figure 5.3D**, and **Table S5.2**). A similar trend was seen for the single mutant strain (**Figure S5.4** and **Table S5.2**). Interestingly, Sac7d-Cpep and Sac7d-Cpep1.1 also inhibited the entry of both single and double mutant forms of HIV (**Figure 5.3D** and **Table 5.1).** This is potentially due to a number of factors, including stabilization and/or slight alteration of the helix, to the extent that high-affinity binding is retained, and contributions from native Sac7d-residues, within the helix or elsewhere.

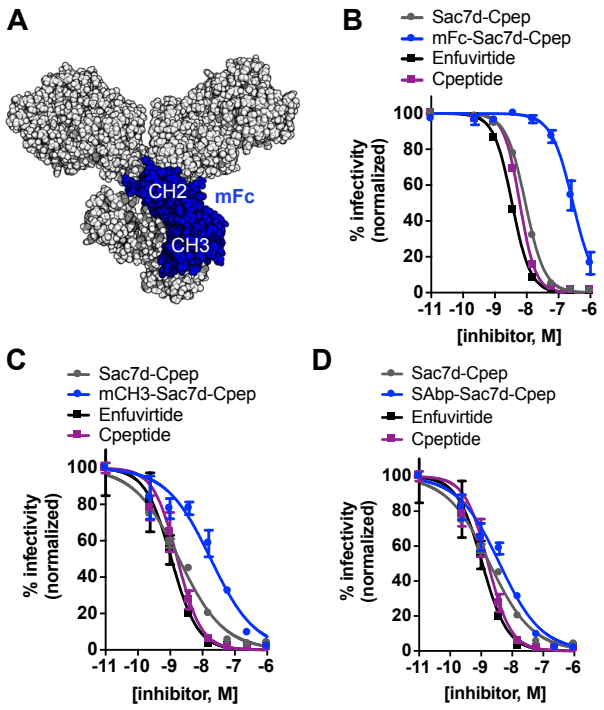


Figure 5.4 (*A*) IgG, with mFc highlighted in blue. mFc consists of mCH2 and mCH3. Fusions either contain complete mFc or the mCH3 domain. (*B*) HIV entry inhibition with mCH3-Sac7d-Cpep, compared to Sac7d-Cpep, Enfuvirtide, and C-peptide. (*C*) HIV entry inhibition with mFc-Sac7d-Cpep, compared to Sac7d-Cpep, Enfuvirtide, and C-peptide. (*D*) HIV entry inhibition with SAbp-Sac7d-Cpep, compared to Sac7d-Cpep, Enfuvirtide, and C-peptide. Error bars represent SEM for three separate experiments. If error bars are not visible, error is smaller.

5.7 Fusions Designed for Serum Stability Potently Inhibit HIV Entry

Small proteins like Sac7d-Cpep, and evolved forms thereof, are typically cleared rapidly. This contrasts to antibodies, which have much longer lifetimes *in vivo* (e.g., IgG half-life is approximately 20 days). Antibody stability is the result of 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).[7-10] This single chain is referred to as monomeric Fc (mFc, **Figure 5.4A**)[11, 12]; a single CH3 domain is referred to as monomeric CH3 (mCH3, **Figure 5.4A**).[13] 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. This biological shell game allows antibodies to evade degradation by proteases. 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*.

On the basis of these earlier reports, we prepared two fusions, mFc-Sac7d-Cpep and mCH3-Sac7d-Cpep, and measured their ability to inhibit HIV entry. Since the version of Sac7d we began with binds IgG, all protein fusions contain a reversion L33T mutation that diminishes affinity for IgG. ELISA experiments indicate no appreciable binding between mFc and mCH3 (**Figure S5.5A**). Consistent with a relationship between protein size and entry inhibition potency, mFc-Sac7d-Cpep (37 kDa) inhibits HIV entry (IC_{50} 282.9 nM, **Figure 5.4B**; **Table 5.1,** entry 18), but is approximately 23-fold less potent than Sac7d-Cpep. mCH3-Sac7d-Cpep (24 kDa) inhibits HIV entry, but the increased size results in 1.5-fold decreased potency (IC_{50} 18.0 nM) compared to Sac7d-Cpep (**Figure 5.4C**; **Table 1**, entry 19). 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 modest increase in size **(Figure S5.5B-C** and **Table S5.2**).

Fusion to a polypeptide that binds serum albumin has also been reported as a strategy to improve *in vivo* stability of biologics.[14] We prepared a fusion consisting of a 20-amino acid serum albumin binding peptide (SAbp) to Sac7d-Cpep (SAbp-Sac7d-Cpep). SAbp-Sac7d-Cpep potently inhibits HIV entry (IC_{50} 3.9 nM; **Figure 5D**; **Table 5.1**, entry 20). We also created a SAbp fusion to the C-terminus of Sac7d-Cpep. To physically separate the two domains, as well as increase the likelihood of helix stabilization, we inserted a helix-forming linker

(AEAAKEAAKA)[15] between Sac7d-Cpep and SAbp. This fusion protein (Sac7d-Cpep-hfl-SAbp) showed similar HIV entry inhibition (IC_{50} 9.7 nM, Figure S6D).

5.8 Conclusion

Here, we report a new, minimalist, helix-grafted display protein (Sac7d-Cpep), that potently inhibits HIV entry in a live virus assay (IC_{50} 1.9-12.4 nM). Yeast display evolution of both grafted and native residues on the solvent-exposed face of the helix led to a library 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 HIV-1 with clinically relevant mutations, leading to Enfuvirtide-resistance. Helix-grafted HIV-1 entry inhibitors fused to either mCH3, mFc, or SAbp, designed to endow serum stability and extended *in vivo* residence, retain the ability to inhibit HIV-1 entry in a live virus assay; however, we do observe a correlation between the size of the fusion and potency of inhibition – larger fusions were less potent. Given the amino acid diversity of evolved biologics we report, 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. We also speculate that a cocktail of these proteins might present a challenge for the emergence of resistance, since it is unlikely that HIV-1 is able to rapidly evolve resistance against the entire therapeutic cocktail.

5.9 Future Directions

Since the mechanism of membrane fusion is similar for enveloped viruses, helix-grafted display could be applied to their corresponding α-helical proteins to find potent inhibitors.

Attaching other viral C- peptides to these PH domains could have varying solubility. PLECKSTRIN, HOMER, ELMO, and Sac7d would be excellent potential scaffolds since they express well as their wild-type. Alternatively there are many other therapeutically relevant protein-protein interactions (PPIs) that involve a helix and a helical binding cleft beyond viral membrane fusion Further investigations could show which PH domains are best for their corresponding target. ELMO was amendable to mutations on the N and C terminus, which could be advantageous given the target. Sac7d's small size is a large benefit for targeting sterically constrained PPIs. Given these strong initial results, the helix-grafted display method shows promise as a platform for helix stabilization and other targets should further investigated to show the broad utility of this platform.

5.10 Methods

Protein Expression and Purification Genes were cloned into pET using BamHI and KpnI, downstream of a His_6 tag and transformed into BL21s (DE3). Cells were grown in 0.5 L LB 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) 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 (4303 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. 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 2XYT and induced at 20 °C for 15-20 hours for maximal yield.

Circular Dichroism Proteins were purified as described. 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.

Cell lysate ELISA Sac7d, Sac7d-Cpep, and C-peptide were cloned into MCS1 of pETDuet with FLAG tags using NcoI and NotI. The 5-Helix with a C-terminal His_6 tag was cloned into MCS2 of pETDuet 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-FLAG 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 Assays 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.

Resolublization of 5-helix Inclusion Bodies 5-helix-His_6 was cloned into a modified pETDuet vector using NdeI and KpnI and transformed into BL21 cells. Cells were induced to express 5-helix-His_6 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-His_6 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. His_6-BirA was cloned into pET using NcoI and KpnI and purified as described above in PBS. 300 mL of Avitagged™ 5-Helix-His_6 protein at 38 μM was incubated with 6 μL of His_6-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 His_6-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-His_6 (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 Na2HPO4, 8.6 g/L NaH2PO4 • H2O, 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 $1.072 \times 10^7 - 2.19 \times 10^7$ 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 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. 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 a-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 Table S2 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 (SAPE) 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.

5.11 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 Cpep

LASEQKLISEEDLGS

C-peptide

WMEWDREINNYTSLIHSLIEESQNQQEKNEQELL

SAbp

QRHPEDICLPRWGCLWGDDD

mCH3

GQCREPQVYTSPPSRDELTKNQVSLRCHVKGFYPSDIAVEWESNGQPENNYKTTKPVLD
SDGSFRLASYLTVDKSRWQQGNVFSCSVMHECLHNHYTQKSLSLSPGK

mFc

APELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKT
KPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQ
VYTKPPSRDELTKNQVSLSCLVKGFYPSDIAVEWESNGQPENNYKTTVPVLDSDGSFRL
ASYLTVDKSRWQQGNVFSCSVMHEALHNHYTQKSLSLSPGK*

5-Helix-His$_6$

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
VKVKFLLNGEEKEVDTSKIRDVSRQGKNVKFLYNDNGKYGAGNVDEKDA
PKELLWMLWDAETNKVMSLLHSLIEESQNQQEKNEQELL

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. Chen, W.; Zhu, Z.; Feng, Y.; Dimitrov, D. S., Human domain antibodies to conserved sterically restricted regions on gp120 as exceptionally potent cross-reactive HIV-1 neutralizers. *Proc Natl Acad Sci U S A* **2008,** *105* (44), 17121-6.

2. Behar, G.; Bellinzoni, M.; Maillasson, M.; Paillard-Laurance, L.; Alzari, P. M.; He, X.; Mouratou, B.; Pecorari, F., Tolerance of the archaeal Sac7d scaffold protein to alternative library designs: characterization of anti-immunoglobulin G Affitins. *Protein Eng Des Sel* **2013,** *26* (4), 267-75.

3. 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* (20), 1945-1950.

4. 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. *Bioorg Med Chem* **2018,** *26* (6), 1220-1224.

5. Hache, G.; Shindo, K.; Albin, J. S.; Harris, R. S., Evolution of HIV-1 isolates that use a novel Vif-independent mechanism to resist restriction by human APOBEC3G. *Curr Biol* **2008,** *18* (11), 819-24.

6. He, Y.; Cheng, J.; Lu, H.; Li, J.; Hu, J.; Qi, Z.; Liu, Z.; Jiang, S.; Dai, Q., Potent HIV fusion inhibitors against Enfuvirtide-resistant HIV-1 strains. *Proc Natl Acad Sci U S A* **2008,** *105* (42), 16332-7.

7. Strohl, W. R., Fusion Proteins for Half-Life Extension of Biologics as a Strategy to Make Biobetters. *BioDrugs* **2015,** *29* (4), 215-39.

8. Qiu, Y.; Lv, W.; Xu, M.; Xu, Y., Single chain antibody fragments with pH dependent binding to FcRn enabled prolonged circulation of therapeutic peptide in vivo. *J Control Release* **2016,** *229*, 37-47.

9. Sockolosky, J. T.; Szoka, F. C., The neonatal Fc receptor, FcRn, as a target for drug delivery and therapy. *Adv Drug Deliv Rev* **2015,** *91*, 109-24.

10. Rath, T.; Baker, K.; Dumont, J. A.; Peters, R. T.; Jiang, H.; Qiao, S. W.; Lencer, W. I.; Pierce, G. F.; Blumberg, R. S., Fc-fusion proteins and FcRn: structural insights for longer-lasting and more effective therapeutics. *Crit Rev Biotechnol* **2015,** *35* (2), 235-54.

11. Ying, T.; Feng, Y.; Wang, Y.; Chen, W.; Dimitrov, D. S., Monomeric IgG1 Fc molecules displaying unique Fc receptor interactions that are exploitable to treat inflammation-mediated diseases. *MAbs* **2014,** *6* (5), 1201-10.

12. Ying, T.; Chen, W.; Gong, R.; Feng, Y.; Dimitrov, D. S., Soluble monomeric IgG1 Fc. *J Biol Chem* **2012,** *287* (23), 19399-408.

13. Ying, T.; Chen, W.; Feng, Y.; Wang, Y.; Gong, R.; Dimitrov, D. S., Engineered soluble monomeric IgG1 CH3 domain: generation, mechanisms of function, and implications for design of biological therapeutics. *J Biol Chem* **2013,** *288* (35), 25154-64.

14. Dennis, M. S.; Zhang, M.; Meng, Y. G.; Kadkhodayan, M.; Kirchhofer, D.; Combs, D.; Damico, L. A., Albumin binding as a general strategy for improving the pharmacokinetics of proteins. *J Biol Chem* **2002,** *277* (38), 35035-43.

15. Chen, X.; Zaro, J. L.; Shen, W. C., Fusion protein linkers: property, design and functionality. *Adv Drug Deliv Rev* **2013,** *65* (10), 1357-69.

APPENDIX A

Chapter 2 Supplemental Data

expected mass (GLUE-Cpep) = 17,069 Da [M+1]
observed mass = 17,069 Da

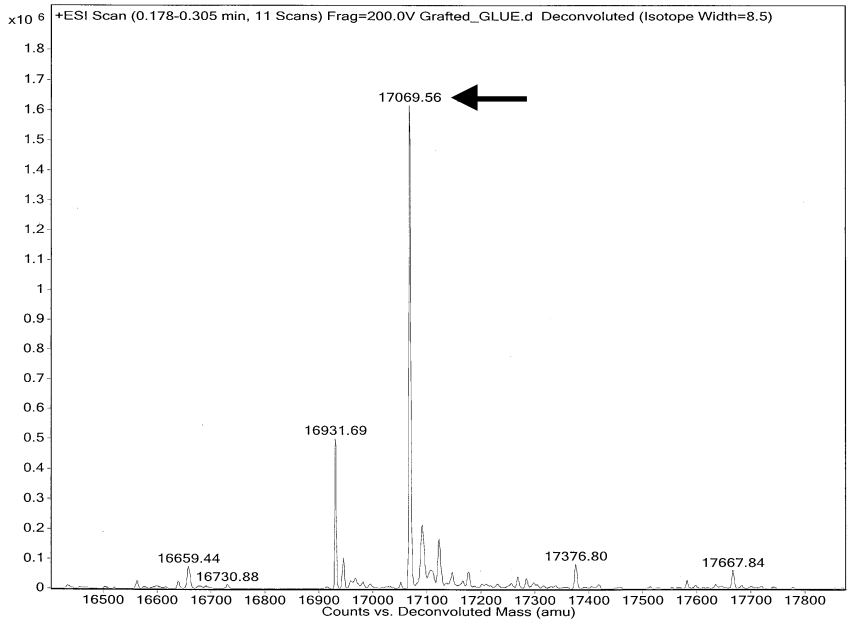


Figure 2.1: Mass of protein (GLUE-Cpep) that is co-purified with 5-helix-His

APPENDIX B

Chapter 3 Supplemental Data

Table S3.1- 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

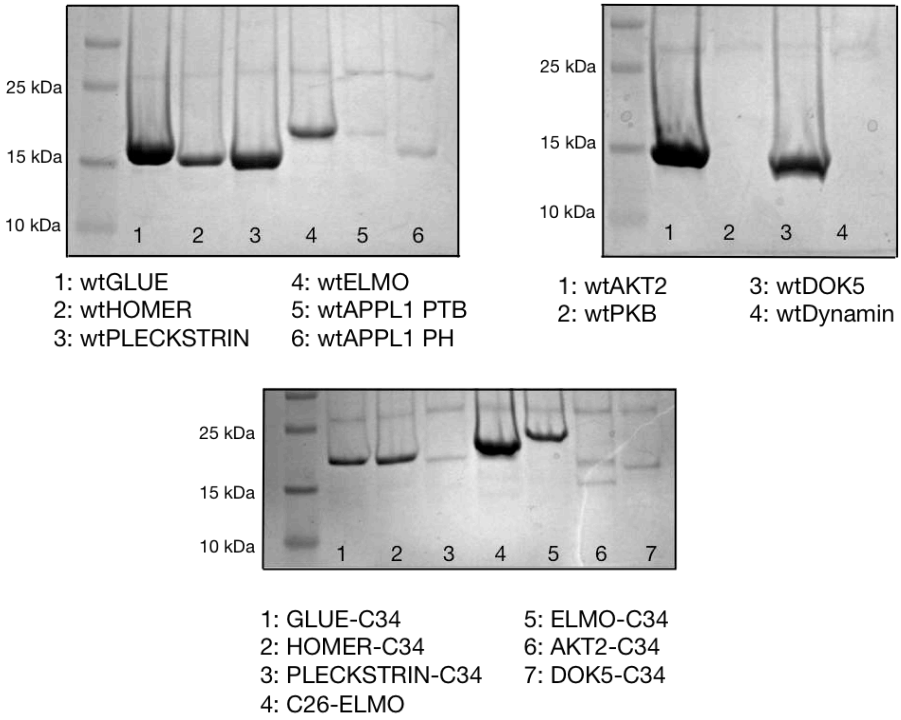


Figure S3.1 SDS page gel and coomassie staining of purified proteins. Top gel show purified wild-type PH domains and bottom gel shows helix grafted proteins

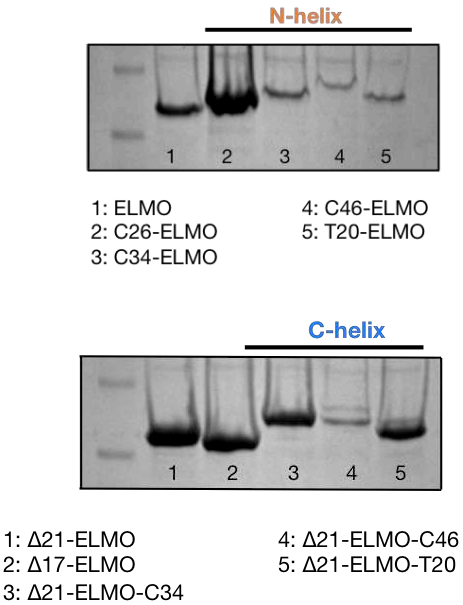


Figure S3.2 SDS page gel and coomassie staining of helix grafted ELMO proteins

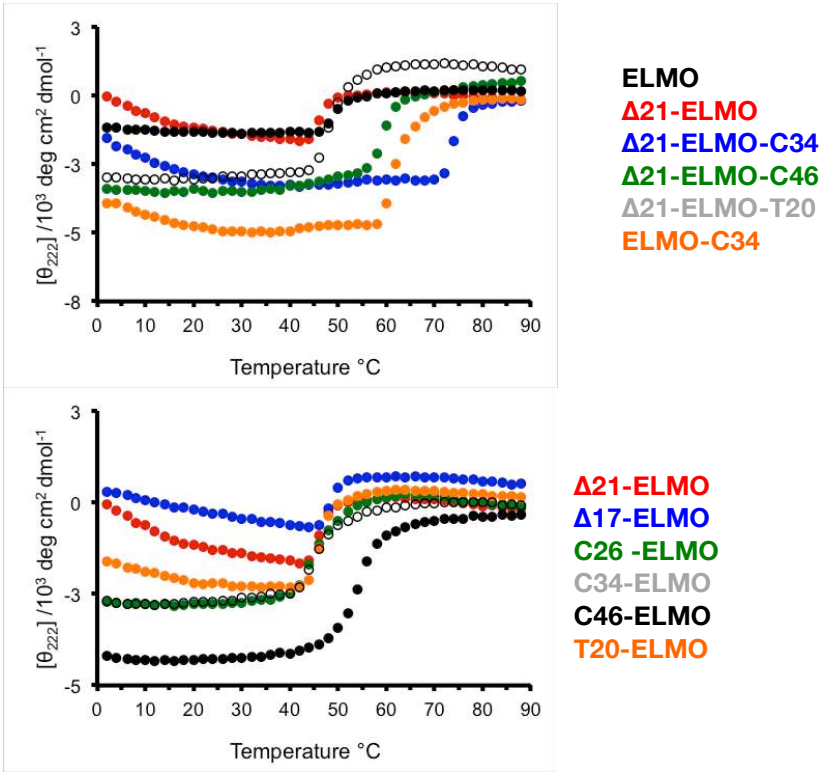


Figure S3.3 Circular dichroism data (222 nm) showing temperature-dependent melting of N and C term helix grafted ELMO proteins

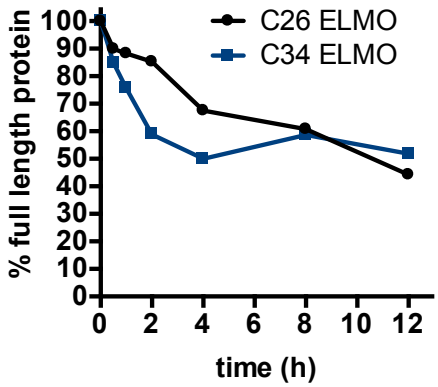


Figure S3.4 % full length helix grafted protein after 12 hour incubation with human serum.

APPENDIX C

Chapter 4 Supplemental Data

Table S4.1: Experimental conditions for yeast display

	[biotin-5helix, nM]	Yeast Screened	Yeast Sorted	% positive
Round 1	500	~7×10^7	~2.8×10^5	0.40%
Round 2	100	~5×10^6	~5×10^5	1.00%

APPENDIX D

Chapter 5 Supplemental Data

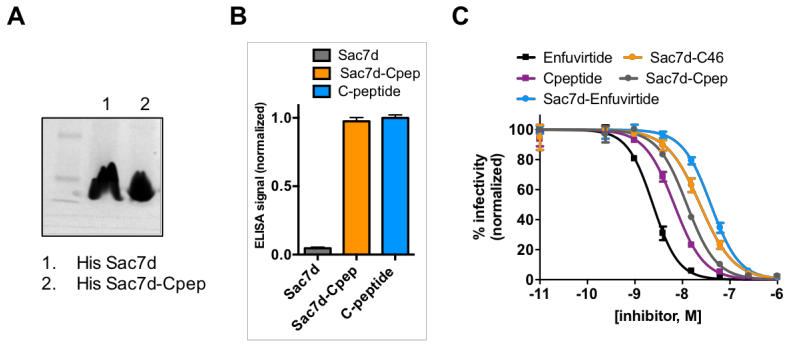


Figure S5.1: Helix Grafted Sac7d-Cpep (*A*) Protein gel of soluble his6-tagged Sac7d and Sac7d-Cpep (*B*) ELISA data from E. coli cell lysate that encodes 5-Helix-His along with Sac7d, Sac7d-Cpep or C-peptide. Proteins were bound to a Ni-coated plate, washed, and incubated with an anti-FLAG antibody to measure binding. (*C*) Helix grafted Sac7d-C46 and Sac7d-Enfuviritide suppresses HIV-1 entry with similar potency to Sac7d-Cpep, C-peptide and Enfuviritide in a live virus assay. 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. Data was plotted log[inhibitor], M.

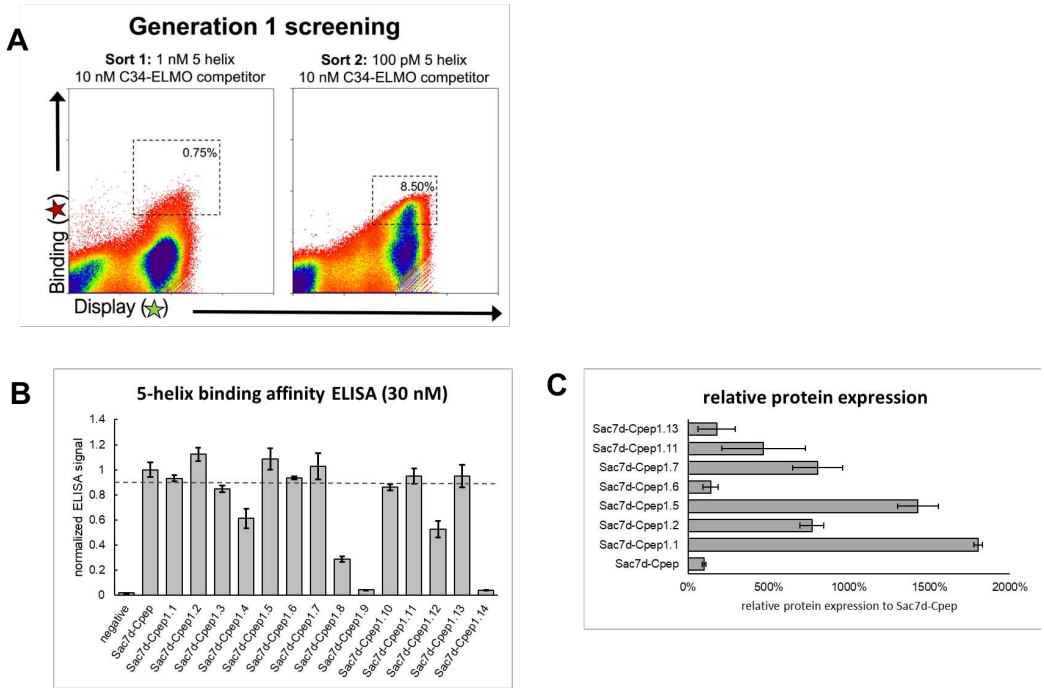


Figure S5.2: Generation 1 Screening and Analysis (*A*) flow cytometry data of generation 1 saturation mutagenesis protein library. Sort 1 contained 0.75 % of the yeast population displaying Sac7d-Cpep based library following incubation with FITC-labelled anti-myc antibody and 1 nM biotin-5-helix / streptavidin-

phycoerythrin, and 10 nM of Cpep-ELMO competitor. Sort 2 contained 8.50 % of the amplified yeast population from the first sort displaying enriched Sac7d-Cpep library following incubation with FITC-labelled anti-myc antibody and 100 pM biotin-5-helix / streptavidin-phycoerythrin, and 10 nM of Cpep-ELMO competitor. (*B*) Affinity for 5-helix recognition was assessed by *in vitro* ELISA at 30 nM and lowest affinity binders were systematically removed from the pool of evolved proteins (less than 90 % of ELISA signal relative to Sac7d-Cpep). (*C*) Protein expression analysis relative to Sac7d-Cpep displays improvement of expression in *E. coli* for all remaining generation 1 evolved protein variants. [16]

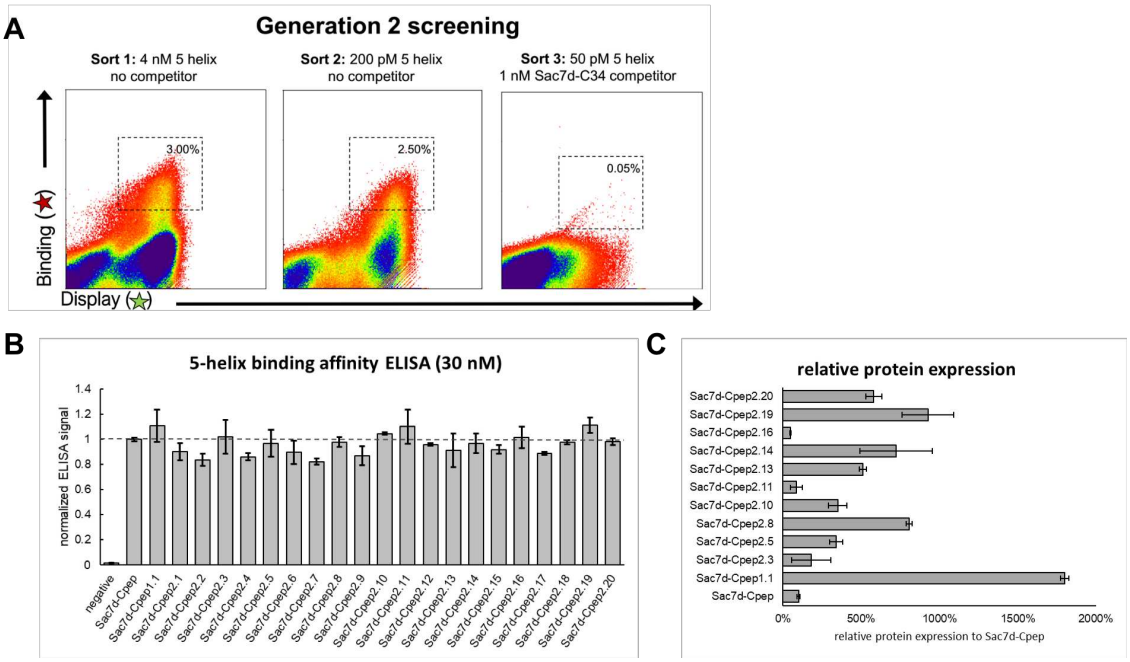


Figure S5.3: Generation 2 Screening and Analysis (*A*) flow cytometry data of generation 2 saturation mutagenesis protein library. Sort 1 contained 3.00 % of the yeast population displaying Sac7d-Cpep based library following incubation with FITC-labelled anti-myc antibody and 4 nM biotin-5-helix / streptavidin-phycoerythrin without competitor. Sort 2 contained 2.50 % of the amplified yeast population from the first sort displaying enriched Sac7d-Cpep library following incubation with FITC-labelled anti-myc antibody and 200 pM biotin-5-helix / streptavidin-phycoerythrin without competitor. Sort 3 contained 0.05 % of the amplified yeast population from the second sort displaying enriched Sac7d-Cpep library following incubation with FITC-labelled anti-myc antibody and 50 pM biotin-5-helix / streptavidin-phycoerythrin, and 1 nM of Sac7d-Cpep competitor. (*B*) Affinity for 5-helix recognition was assessed by *in vitro* ELISA at 30 nM and lowest affinity binders were systematically removed from the pool of evolved proteins (less than 100 % of ELISA signal relative to Sac7d-Cpep). (*C*) Protein expression analysis relative to Sac7d-Cpep displays improvement of expression in *E. coli* for the majority of remaining generation 2 evolved protein variants.

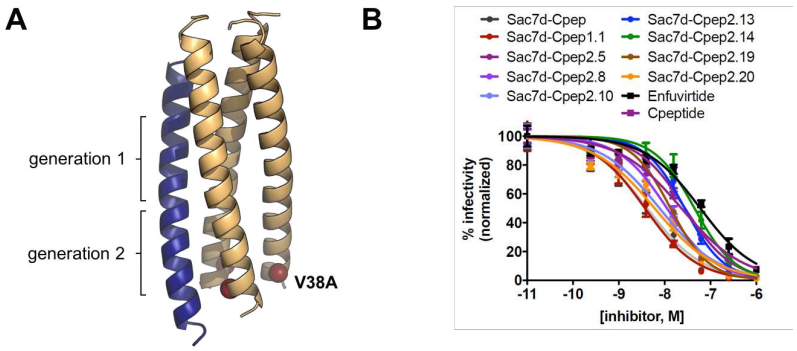


Figure S5.4: Polyclonal Potential for Enfuvirtide Resistance (*A*) Crystal structure of the NHR trimeric prefusogenic conformation with one hydrophobic groove occupied by a Cpeptide (PDB:1AIK). The labelled red spheres represent a common mutation found in Enfuvirtide-resistant (Cpeptide sensitive) HIV-1 strains and its location relative to the variable regions of the evolved Sac7d-Cpep proteins. (*B*) Entry inhibition data for the second generation of evolved proteins was assessed by a live virus assay, and all the grafted proteins showed improvement in potency compared to Enfuvirtide (IC_{50} less than 60.0 +/- 11.6) and six of the grafted proteins showed improvement in potency compared to C-peptide (IC_{50} less than 23.2 +/- 4.1). Data was plotted log[inhibitor], M.

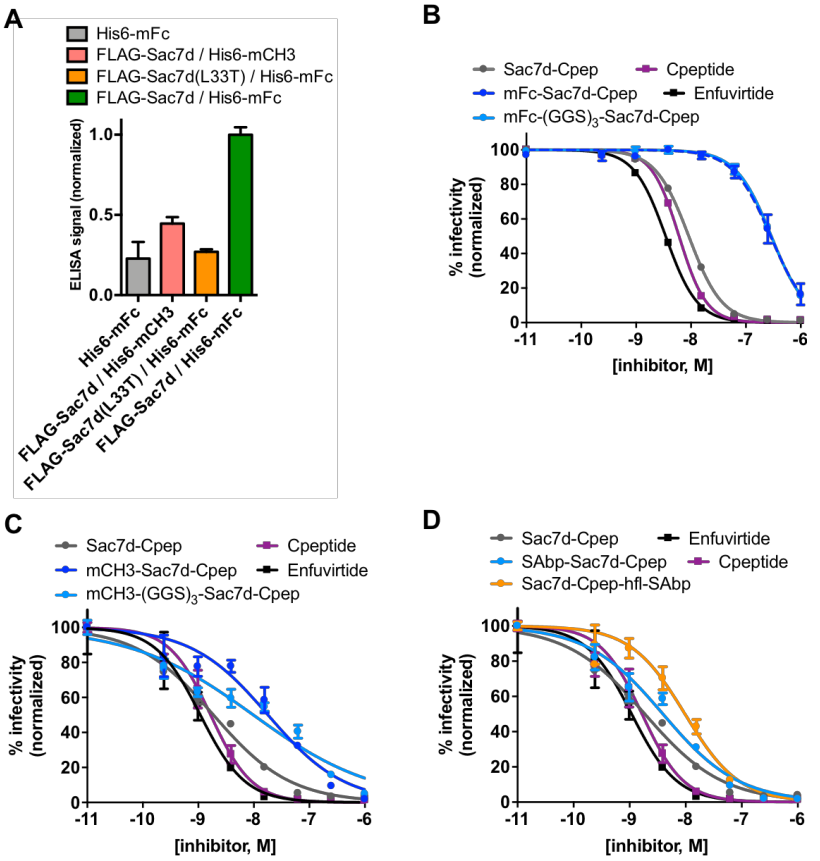


Figure S5.5. Linker Optimization Infectivity (*A*) ELISA data from E. coli cell lysate that encodes either His-mCH3 or mFc in MCS2 along with FLAG Sac7d or FLAG Sac7d L33T (off) in MCS1. Proteins were bound to a Ni-coated plate, washed, and incubated with an anti-FLAG antibody to measure binding. (*B*) Infectivity mFc Sac7d-Cpep with and without a $(GGS)_3$ linker. (*C*) Infectivity mCH3 Sac7d-Cpep with and without a $(GGS)_3$ linker. (*D*) Infectivity of N- and C- terminal SAbp

fusions. The N terminal fusion does not contain a linker between the fusion peptide and Sac7d-Cpep. The C terminal fusion contains a helix-forming linker in-between C-peptide and SAbp to help maintain structure. All infectivity data was plotted log[inhibitor], M.

Table S5.1. Library Screening Schematic

	[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 x 10^7	~5.25 x 10^5	0.75%
Generation 1: Round 2	0.10	10.00	0.00	~1 x 10^7	~1.0 x 10^6	1.00%
Generation 2: Round 1	4.00	0.00	0.00	~8 x 10^7	~2.4 x 10^6	3.00%
Generation 2: Round 2	0.20	0.00	0.00	~2 x 10^7	~5.0 x 10^5	2.50%
Generation 2: Round 3	0.05	0.00	1.00	~6 x 10^6	~3.0 x 10^4	0.05%

Table S5.2: Cumulative Infectivity IC_{50} values

				generation 1					generation 2					IC50 (nM)	V38A IC50 (nM)	V38A/N42D IC50 (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	-	-

LIST OF ABBREVIATIONS

AIDS	aquired immunodeficiency syndrome
Ala/A	alanine
Arg/R	arginine
Asn/N	asparagine
Asp/D	aspartic acid
BSA	bovine serum albumin
CD	circular dichroism
Cfu	colony forming unit
Cys/C	cysteine
Da	dalton
DNA	deoxyribonucleic acid
DTT	dithiothreitol
ELISA	enzyme-linked immunosorbent assay
ELMO	engulfment and cell motility
FACS	fluorescence-activated cell sorting
FDA	food and drug administration
FITC	fluorescein isothiocyanate
GFP	green fluorescent protein
Gln/Q	glutamine
Glu/E	glutamic acid

GLUE	gram like ubiquitin binding in EAP45
Gly/G	glycine
HA	hemagglutinin
His/H	histidine
HIV	human immunodeficiency virus
HRP	horseradish peroxidase
Ile/I	isoleucine
IPTG	β-D-1-thiogalactopyranoside
K_D	dissociation Constant
LC-MS	liquid chromatography mass spectrometry
Leu/L	leucine
Lys/K	lysine
MCS	multiple cloning site
Met/M	methionine
NHR	N-heptad repeat
NTA	nitrilotriacetic acid
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PDB	protein data bank
PE	phycoerythrin
PH Domain	pleckstrin homology domain

Phe/F	phenylalanine
PPI	protein-protein interaction
Pro/P	proline
RT	room temperature
RPM	revolutions per minute
SAbp	serum albumin binding peptide
SAPE	streptavidin-R-phycoerythrin
SDS	sodium dodecyl sulfate
SEM	standard of the mean
Ser/S	serine
sfGFP	superfolder green fluorescent protein
spGFP	superpositive green fluorescent protein
Thr/T	threonine
TOF	time of flight
TMB	3, 3′, 5, 5′-tetramethylbenzidine
Tris	tris(hydroxymethyl) aminomethane
Trp/W	tryptophan
Tyr/Y	tyrosine
Val/V	valine