REFLECTING ON LIFE: ADVENTURES IN MIRROR-IMAGE BIOLOGY

by

Matthew T Weinstock

A dissertation submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Biochemistry

The University of Utah

May 2014

The University of Utah Graduate School

STATEMENT OF DISSERTATION APPROVAL

The following faculty members served as the supervisory committee chair and members for the dissertation of **Matthew T Weinstock**.

Dates at right indicate the members' approval of the dissertation.

Michael S. Kay	, Chair	**03/03/14** Date Approved
Wesley I. Sundquist	, Member	**03/03/14** Date Approved
Christopher P. Hill	, Member	**03/03/14** Date Approved
Eric W. Schmidt	, Member	 Date Approved
Dennis R. Winge	, Member	**03/03/14** Date Approved

The dissertation has also been approved by **Christopher P. Hill**, Chair of the Department/School/College of **Biochemistry** and by David B. Kieda, Dean of The Graduate School.

ABSTRACT

The presence of stereogenic atoms in many biomolecules endows life with the property of chirality, making it possible that a mirror-image biological universe could exist. This body of work explores avenues to, and uses for, mirror-image biological systems. Particular attention is placed on the applications of mirror-image biology in the areas of drug discovery and synthetic biology.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

Were I to include everyone worthy of mention here, the acknowledgements section would exceed the length of the actual text of this thesis; nevertheless, I would like to take the time to acknowledge the outstanding contributions of a few.

First I would like to express my gratitude to my best friend and loving wife, Natasha, who has been a great encouragement throughout our friendship and was a great sport during my time in graduate school despite my scattered brain, frazzled nerves, long hours, and low salary. I am indebted to her for the sacrifices she has made to make graduate school work with our family. I would also like to acknowledge my two children, Jacob and Joshua, who consistently brightened my days, even though they kept me up many nights.

I am very grateful to my parents, Timothy and Sandra Weinstock, who provided me with a safe, loving environment where I was blessed with all sorts of opportunity. They have been and continue to be extremely supportive of me in my endeavors.

Dr. Thomas Richmond served as an advisor to my undergraduate education at the University of Utah, and granted me the privilege of working in his lab during my undergraduate education. He was also the one who called my attention to the Biological Chemistry program at the University of Utah, where I ended up pursuing my graduate degree.

Dr. Michael Kay, in whose lab I pursued my graduate education, deserves special acknowledgement. He has proven to be an exceptional mentor, teacher, and friend. I am especially thankful for his supporting me in my own scientific interests, encouraging me to attend conferences that I find interesting, sharing articles he knew I would appreciate, and providing all sorts of helpful life advice. He truly has a heart for his students and is interested in helping them develop as whole individuals. His influence will remain with me throughout my life.

I have enjoyed the company of many great colleagues in the Kay lab, including Brett Welch, Joseph Redman, Nicholas Francis, Tracy Clinton, Ethan Howell, Sarah Apple, and Maya Pandya. Special thanks are due to Michael Jacobsen, who collaborated with me on various projects in the Kay lab; Rob Marquardt, who served for a while as lab manager and taught me everything I know about mass spectrometry; and Debbie Eckert, who oversaw my initial rotation in the Kay lab, was kind enough to provide critical readings of my various attempts at scientific writing, and from whom I have learned an immense amount of science.

I am very thankful to the members of my graduate committee, Wesley I. Sundquist, Christopher P. Hill, Eric Schmidt, and Dennis Winge. Besides being granted the privilege to rotate through the labs of three out of four members, they all provided critical input as I progressed throughout my graduate career. I greatly appreciated their investment in my growth as a scientist.

I would also like to acknowledge those in the broader Biochemistry community at

the University of Utah, particularly Frank Whitby, who I always enjoyed talking with during our late night crystallographic data collection sessions; Costa Georgopolous and Debbie Ang, who were incredibly gracious with their time and reagents in providing assistance on one of my projects, and who were always encouraging and kind; and Dana Carroll, who kindly offered to bring me along to a Keystone symposium on genome engineering/synthetic biology that he organized as his conference assistant. It was at this conference that I made important connections leading to my postgraduate position at Synthetic Genomics, Inc.

CHAPTER 1

PROLEGOMENON

The unifying theme of this thesis is the use of mirror-image proteins for drug discovery (via mirror-image phage display) and synthetic biology applications.

Chapter 2 serves as a general introduction by putting our mirror-image phage display efforts in the larger context of peptidomimetic drug discovery, highlighting the advantages of D-peptides as therapeutics, and describing methods for their discovery via genetically-encoded, library-based screening techniques. Chapter 2 also provides an introduction to strategies for the synthesis of mirror-image proteins for use as display targets and proposes some exciting future directions for the use of mirror-image proteins in the field of synthetic biology.

Chapter 3 provides a more detailed look at the D-peptide discovery process by describing the design of a third-generation D-peptide HIV-1 entry inhibitor. This was achieved through the optimization of a second-generation D-peptide entry inhibitor (PIE7) by structure-guided phage display and improvement of the geometry of PEG crosslinkers used to oligomerize the peptide into dimers and trimers. The work culminated in the development of third-generation D-peptide (PIE12-trimer) that exhibits dramatically higher affinity for its binding site, improved potency, and a higher barrier to resistance.

Chapter 4 sets forth the current state of a project aiming to elucidate resistance

mechanisms employed by HIV against D-peptide entry inhibitors. It describes our methodology to comprehensively characterize the genetic makeup of resistant viral populations through the use of both traditional Sanger and deep sequencing technologies, and describes the biochemical characterization of a resistance mechanism against the D-peptides PIE12-trimer and cholesterol-conjugated PIE12-trimer. This chapter concludes with a discussion of strategies to overcome resistance.

Chapter 5 describes the extension of our D-peptide discovery process to new targets, describing initial work towards the development of D-peptide entry inhibitors against Ebola virus. Although Ebola has a similar entry mechanism to HIV, Ebola viral fusion occurs in the endosomal compartment versus at the cell surface (HIV). This imposes the additional challenge of targeting peptides to the endosome. Chapter 5 presents work on the design, synthesis, and biophysical characterization of Ebola targets and preliminary target validation via phage display methods.

Chapter 6 addresses the challenge of mirror-image protein synthesis. Because mirror-image drug discovery and mirror-image synthetic biology are currently dependent upon our ability to successfully chemically assemble mirror-image proteins, robust methods for their construction are a priority. Work presented in this chapter pushes the limits of current synthetic protein assembly methodologies in the total chemical synthesis of a 312 residue protein (dapA) in both L- and D- chiralities (the largest synthetic proteins in either chirality to date). dapA was chosen for this study because of its dependence upon the GroEL/ES chaperone system for folding both *in vitro* and *in vivo* under physiological conditions. After we demonstrated that our synthetic constructs were active by identifying nonphysiological buffer conditions that promoted folding, we used

the L- and D- enantiomers of dapA to probe for chiral specificity in the GroEL/ES protein-folding machinery.

CHAPTER 2

PROTEASE-RESISTANT PEPTIDE DESIGN – EMPOWERING NATURE'S FRAGILE WARRIORS AGAINST HIV

Matthew T. Weinstock*, J. Nicholas Francis*, Joseph S. Redman, Michael S. Kay

* These authors contributed equally to this work

Reprinted with permission from Peptide Science,

Vol. 95, Issue 5, pp. 431-442

Invited Review
Protease-Resistant Peptide Design—Empowering Nature's Fragile Warriors Against HIV

Matthew T. Weinstock, J. Nicholas Francis, Joseph S. Redman, Michael S. Kay
Department of Biochemistry, University of Utah School of Medicine, Salt Lake City, UT 84112-5650

Received 23 December 2011; revised 5 March 2012; accepted 4 April 2012
Published online 14 April 2012 in Wiley Online Library (wileyonlinelibrary.com). DOI 10.1002/bip.22073

ABSTRACT:

Peptides have great potential as therapeutic agents, but their use is often limited by susceptibility to proteolysis and their resulting in vivo *fragility. In this review, we focus on peptidomimetic approaches to produce protease-resistant peptides with the potential for greatly improved clinical utility. We focus on the use of mirror-image (D-peptide) and ß-peptides as two leading approaches with distinct design principles and challenges. Application to the important and difficult problem of inhibiting HIV entry illustrates the current state-of-the-art in peptidomimetic technologies. We also summarize future directions for this field and highlight remaining obstacles to widespread use of protease-resistant peptides. © 2012 Wiley Periodicals, Inc. Biopolymers (Pept Sci) 98: 431–442, 2012.*

Keywords: peptidomimetics; HIV entry; peptide design

This article was originally published online as an accepted preprint. The "Published Online" date corresponds to the preprint version. You can request a copy of the preprint by emailing the Biopolymers editorial office at biopolymers@wiley. com

Correspondence to: Michael S. Kay, Department of Biochemistry, University of Utah School of Medicine, 15 N Medical Drive East Rm 4100, Salt Lake City, UT 84112-5650, USA; e-mail: kay@biochem.utah.edu
Matthew T. Weinstock and J. Nicholas Francis contributed equally to this work.
Contract grant sponsor: NIH (to M.S.K.)
Contract grant number: AI076168
Contract grant sponsor: NIH Microbial Pathogenesis Predoctoral Training Grant (to J.N.F.)
Contract grant number: AI055434

INTRODUCTION

In drug discovery and development, peptide therapeutics have many advantages. Their polymeric nature makes synthesis straightforward, especially when compared with the synthetic schemes typically utilized for small molecules. Peptides are generally easier and less expensive to produce than recombinant proteins. Peptide therapeutics can also be more specific (and less toxic) than small molecules and excel at the challenging problem of disrupting large protein–protein interaction interfaces (i.e., "undruggable" targets). Due to advancements in genomics and proteomics, a plethora of natural peptide ligand sequences for important drug targets are available and provide a sensible starting point for the rational development of therapeutic compounds. In addition, a host of mature and emerging library-based screening techniques provides a means to rapidly discover novel peptide sequences with specific binding properties.

Despite these enticing advantages, a major problem limiting development of peptide therapeutics is their proteolytic sensitivity and associated delivery challenges. Synthetic therapeutic peptides are typically relatively unstructured and are therefore rapidly degraded *in vivo*, often with half-lives on the order of minutes.[1] Proteolysis commonly occurs in the GI lumen, intestinal brush border, enterocytes, hepatocytes, antigen-presenting cells, and plasma. Because of this *in vivo* fragility, oral delivery is generally not possible, necessitating frequent dosing by injection. Even when delivered parenterally, degradation in the blood combined with rapid renal filtration often results in drugs that are expensive, inconvenient, and unpleasant to administer.

Protease-resistant peptides would address many of these limitations. One of the most promising approaches is to modify the chemical structure of the peptide backbone (peptidomimetics).[2] Modifications that have been shown to substantially decrease proteolysis include N-methylation, ester

linkages (α-hydroxy acids), insertion of additional methylene groups into the backbone (β-amino acids, γ-amino acids etc.), and the use of D-amino acids. More significant changes to the peptide backbone include peptoids, azapeptides, oligoureas, arylamides, and oligohydrazides.[2–4]

In this review, we describe how modified peptide backbones can be used to design protease-resistant inhibitors with a special focus on the high-priority problem of designing protease-resistant HIV entry inhibitors. Although these modified backbones effectively address protease sensitivity, each is associated with a set of design challenges using rational design or library screening techniques. This review will not cover traditional strategies to reduce protease sensitivity, e.g., peptide capping, sequence alteration at susceptible sites, cyclization, or stapling, which have been extensively reviewed elsewhere.[5]

INHIBITING HIV ENTRY

An estimated 34 million people worldwide are infected with HIV, the causative agent of AIDS, resulting in nearly 2 million deaths per year and over 25 million cumulative deaths (UNAIDS). Dramatic progress has been made in reducing mortality since the inception of antiretroviral therapy against HIV enzymes reverse transcriptase, protease, and recently integrase. However, the relentless development of drug resistance necessitates ongoing development of therapeutics that target other stages in the viral lifecycle. In particular, there have been extensive efforts to develop potent, broadly active, and economical entry inhibitors for the prevention and treatment of HIV/AIDS.[6]

The current HIV entry pathway model is shown in Figure 1. Viral entry into host cells is mediated by the trimeric HIV envelope (Env) glycoprotein. Env contains the noncovalently associated surface gp120 and transmembrane gp41 subunits. gp120's primary function is to interact with cell receptors that mark HIV's preferred target cells (e.g., T-cells and macrophages), while gp41 induces membrane fusion. Host cell interactions are mediated by gp120 through association with the primary cell receptor (CD4) and chemokine coreceptor (either CXCR4 or CCR5, depending on viral tropism). Upon gp120 engagement with cell receptors, a complex series of structural rearrangements in gp120 propagate to gp41, activating it for membrane fusion (reviewed by Ref. 7). At this stage, gp41 forms an extended prehairpin intermediate containing an N-terminal trimeric coiled coil (N-trimer) and C-terminal region (C-peptides) of unknown structure. Fusion is driven by collapse of this intermediate as three helical C-peptides pack anti-parallel to the N-trimer (trimer-of-hairpins formation), drawing the viral and host cell membranes into close proximity.

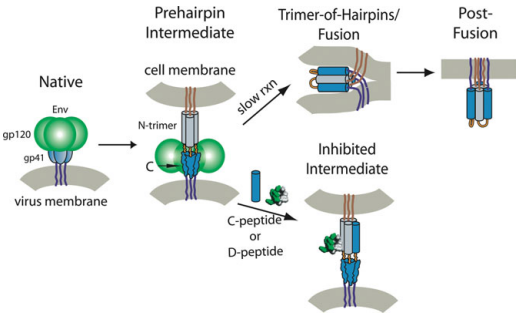


FIGURE 1 HIV entry pathway. HIV Env is composed of surface (gp120, green) and transmembrane (gp41, blue) subunits. Fusion is initiated by binding to CD4 and a chemokine coreceptor, which activates gp41 and induces formation of the prehairpin intermediate. In this intermediate, the gp41 N-terminal region forms a trimeric coiled coil (N-trimer, gray), which is separated from the C-peptide region (dark blue). This intermediate slowly collapses to form a trimer-of-hairpins structure that brings the viral and cell membranes into close apposition, leading to fusion. C-peptide and D-peptide inhibitors bind to the N-trimer, preventing trimer-of-hairpins formation and membrane fusion.

A similar fusion mechanism is utilized by many other enveloped viruses, including influenza, Ebola, and paramyxoviruses.[7]

C-Peptide Inhibitors

This mechanism suggests that peptides derived from the N- and C-peptide regions of gp41 could prevent viral membrane fusion in a dominant-negative manner by preventing trimer-of-hairpins formation. Indeed, both N-and C-peptides inhibit HIV entry.[8–14] The N-trimer/C-peptide interaction is predominantly mediated by conserved interactions between the hydrophobic face of helical C-peptides and a hydrophobic groove formed between helices in the N-trimer. C-peptide inhibitors are more promising drug candidates because of their higher potency and better solubility compared with N-peptide inhibitors.

C-peptide inhibitors were first identified through screens of gp41-derived peptides.[9,11] Fuzeon (Enfuvirtide, T-20) is a 36 amino acid L-peptide taken from the gp41 C-peptide region. Fuzeon inhibits HIV entry with nM potency and reduces viral loads by 2 logs,[15] leading to its approval as the first HIV entry inhibitor in 2003. Unfortunately, Fuzeon's clinical use has been limited by its short half-life. Fuzeon requires injection at very high doses (90 mg, twice daily) to overcome its proteolysis and rapid renal filtration. These practical problems result in a drug that is expensive (~$30,000 per year), can cause painful injection site reactions, and is only approved for patients experiencing treat-

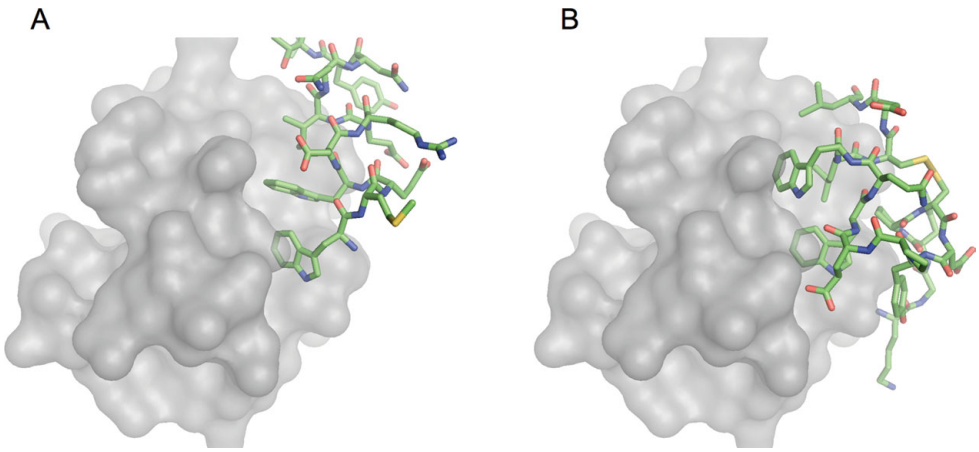


FIGURE 2 One pocket, two binding solutions. The gp41 pocket (from pdb code 3L35) is shown with (A) the natural gp41 C-peptide (pdb code 1AIK) and (B) D-peptide PIE12 (pdb code 3L35). Structures were aligned on the 17 pocket-forming residues from gp41 and rendered using Pymol.

ment failure due to multi-drug resistance ("salvage therapy"). Fuzeon's high dosing requirements and *in vivo* fragility also limit options for less frequent dosing via depot formulation.

The gp41 "Pocket" Region

At the N-trimer's C-terminus lie three symmetry-related deep hydrophobic pockets. Each pocket has a volume of ~400 Å^3 that is filled primarily by three C-peptide residues (Trp628, Trp631, and Ile635)[16,17] (Figure 2). The pocket is a promising inhibitory target because of its critical importance in membrane fusion and very high level of conservation across diverse HIV strains.[16,18] Mutations in the pocket are often not well tolerated due to the requirement for compensatory mutations in the C-peptide region to restore binding. In addition, the pocket region is encoded by the structured RNA region of the Rev-responsive element (RRE), which contains a signal critical for nuclear export of viral RNA.[18] Interestingly, extensive efforts by numerous groups to discover small molecule pocket-binding inhibitors have had limited success, generally producing inhibitors with modest potency and/or significant toxicity.[19–23] Based on this body of work, the gp41 pocket appears to be "undruggable" by small molecule inhibitors, a common problem for extended protein–protein interaction interfaces.

Fuzeon was discovered before the gp41 6-helix bundle crystal structure and does not bind to the gp41 pocket. However, next generation C-peptide inhibitors (e.g., C34, T-1249) do include pocket-binding residues and enjoy superior potencies and resistance profiles.[24–26] The follow-on compound to Fuzeon, T-1249, performed very well in clinical trials, but was not developed further due to unspecified formulation problems, which we speculate includes challenges in the economic synthesis of this 39-residue peptide and a requirement for four 1 mL injections, once per day, as used in a phase I/II trial.[25]

Fuzeon and T-1249 show that a peptide fusion inhibitor can be very effective against HIV, but the impact of such drugs will be limited until the problems of short half-life and high dosing (and the resulting high cost) can be overcome. In this review, we focus on two distinct strategies that have yielded promising protease-resistant peptide fusion inhibitors with the potential to overcome Fuzeon's *in vivo* fragility.

RATIONAL DRUG DESIGN WITH MODIFIED PEPTIDE BACKBONES

While there is much interest in the *de novo* development of peptides with defined structural and functional characteristics, this work is hampered by limitations in currently available modeling strategies. Thus, as illustrated below, most successful rational designs of protease-resistant peptides start from sequence and structural information from existing peptide ligands.

In the realm of rational design of modified peptide therapeutics, β-peptides and mixed α/β-peptides are among the most promising. β-peptides are composed of β-amino acids, which contain an extra backbone methylene group (between

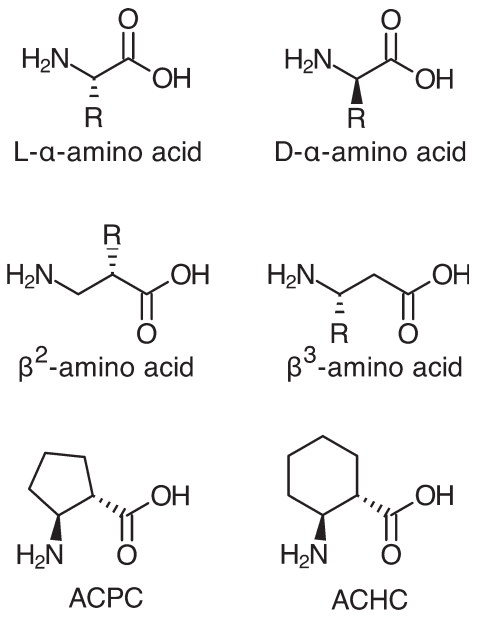


FIGURE 3 Peptidomimetic structures.

the amino and α-carbon, specified as a β^2-amino acid, or between the carboxylate and α-carbon, specified as a β^3-amino acid) (Figure 3). Short β-peptide sequences can adopt robust secondary structures analogous to α-helices formed by α-amino acids. If a natural helical peptide ligand is known, a β-peptide mimic can be generated by the precise placement in three dimensions of key side chains onto a β-peptide scaffold. Two β-peptide scaffolds that have been extensively utilized are the 12-helix and 14-helix, named after the number of atoms between hydrogen bonding groups (these and other β-residue-containing scaffolds are reviewed elsewhere[3,27–30]). The specific structural motif adopted by a particular β-peptide is dictated by the nature of the substituent β-amino acids.[31] β-peptides composed of monosubstituted, acyclic β-amino acids or cyclic six-member ring β-amino acids preferentially adopt the 14-helix structure, while the 12-helix structure is favored by peptides composed of cyclic five-member ring β-amino acids. The helical parameters of the 12- and 14-helices are discussed and compared with α-helices in Refs. 27 and 31.

In a 14-helix composed of β^3-amino acids, side chains at residues *i*, *i*+3, and *i*+6 are presented along the same face of the helix, and are reasonably superimposable with side chains at residues *i*, *i*+4, and *i*+7 of an α-helix.[32] This property can be exploited to display epitopes that mimic an α-helical face and has been applied to the development of low-mid μM HIV entry inhibitors that bind to the gp41 pocket region.[33,34] In an analogous approach, β-peptide inhibitors of HCMV entry were developed using the 12-helix scaffold.[35] To map an α-helix epitope onto the 12-helix, side chains at positions *i*, *i*+4, and *i*+7 on the α-helix are placed at positions *i*, *i*+3, and *i*+5 on the 12-helix. Although acyclic residues diminish 12-helix propensity, they provide the easiest avenue for side chain attachment, so a minimum number of acyclic β^2 or β^3 residues were introduced into the structure at specific points to mimic side chain presentation of the native α-helix. This approach enabled the rapid discovery of inhibitors with modest potency, but its main challenge is the lack of a route forward, by rational design or high-throughput screening, to optimize these initial hits.

A sequence-based approach utilizing mixed α/β-peptides has been applied to develop an HIV entry inhibitor that structurally and functionally mimics C-peptides (∼10 turn α-helix).[36] In this approach, a subset of C-peptide residues were strategically replaced with homologous β^3-amino acids following an ααβααβ pattern, which, despite the additional methylene units, does not significantly alter secondary structure of the helix.[37] On folding, this pattern generates an α-helix-like conformation with a β-residue stripe that runs down the side of the helix distal to the interaction surface, minimizing disruption of the binding interface. On replacing 11 of the 38 residues with β^3-amino acids, the resulting α/β-peptide had >10,000-fold diminished affinity for its binding target relative to the α-peptide counterpart.

As a second step in the design, specific β^3-residues were replaced by cyclic β-residue homologues. The cyclic residues were incorporated to reduce the entropic penalty associated with helix formation due to the inherent torsional flexibility of β^3-residues. β^3 analogues of alanine in the α/β-peptide were replaced with a nonpolar, five-member ring constrained β-residue (ACPC), while β^3 analogues of arginine were replaced with a polar, heterocyclic analogue of ACPC (APC). These replacements improved affinity by ∼400-fold over the peptide with acyclic residues. Although the binding affinity never recovers to that of the original α-peptide ligand, the resulting α/β-peptide was nearly as potent as the α-peptide, but with the added advantage of being 280-fold more resistant to proteolytic degradation by proteinase K. The apparent discrepancy of having diminished binding affinity, yet α-peptide-like potency is likely due to the potency plateau observed for many HIV entry inhibitors (see the discussion of the "resistance capacitor" below).

The original report indicated that the N-terminal Trp-Trp-Ile motif of the α/β-peptide does not engage the C-terminal hydrophobic pocket of gp41, but subsequent crystallographic analysis indicated that that the pocket-binding motif

on the α/β-peptide is indeed able to engage the pocket. The authors suggest that the lack of engagement in the original structure was an artifact caused by crystal packing, and that the newer structure more faithfully portrays the binding of the α/β-peptide (see discussion in supplementary materials of Ref. 38).

GENETICALLY ENCODED LIBRARY-BASED SCREENS

An alternative to rational design is screening of random peptide libraries. These high-throughput methods identify novel peptides with a desired function (typically binding to an immobilized target). Commonly used screening techniques include phage, ribosome, and mRNA display, but these methods all rely on cellular translation machinery and are therefore not yet fully compatible with peptidomimetics in their standard forms. Though there have been many advances and refinements in the field of synthetic peptidomimetic library generation (e.g., split and pool synthesis, physically addressable synthesis by photolithography), these synthetic libraries are typically limited to $<10^6$ members[39] compared with the billion to trillion member libraries that can be generated with genetically encoded libraries.

D-Peptides

D-peptides are entirely composed of D-amino acids, which are mirror-image stereoisomers of the L-amino acids found in naturally occurring L-peptides. D-Peptides are a promising therapeutic platform because they are highly resistant to natural proteases.[40] In elegant work by the Kent group,[41] D-HIV protease was shown to cleave only D-peptide substrates, showing that proteases exhibit highly stereospecific substrate discrimination.

The symmetry relationship between L- and D-peptides can be exploited in mirror-image display techniques[42] in which a mirror-image version of the target molecule is generated by solid-phase synthesis using D-amino acids. Randomized genetically encoded L-peptide libraries are then screened against this D-target. The winning L-peptides are identified by DNA sequencing and then the corresponding D-peptides are synthesized. By symmetry, the D-peptide will have the same activity toward the natural L-target as the L-peptide had against the mirror-image D-target (Figure 4).

A major limitation of mirror-image display is the requirement for chemical synthesis of the D-target. Synthesis of D-peptides is currently done using traditional solid phase peptide synthesis (SPPS).[43] Routine use of SPPS chemistries for the production of peptides is limited to ~50 residues,

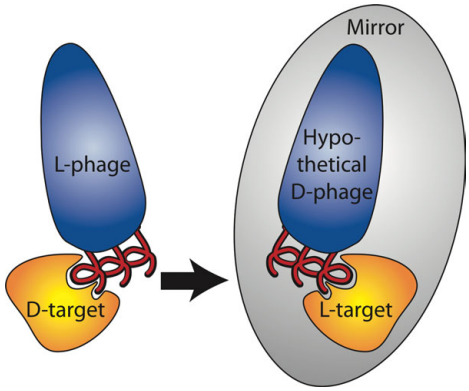


FIGURE 4 Mirror-image phage display. Phage bearing L-peptides are panned against a mirror-image protein (D-target). By symmetry, D-versions of binding peptides will bind to the natural L-targets.

though this limit varies widely depending on the required purity and sequence/structure properties of the peptide in question (e.g., extended beta-strand peptides can aggregate during SPPS). Despite these challenges, syntheses of very long peptides have been reported (e.g., the synthesis of the 140-residue IL-3 protein[44]).

Larger D-peptide targets can be obtained using peptide ligation techniques to link multiple synthesized peptide fragments. A variety of ligation chemistries have been developed (see Ref. 45 for a very thorough review), but the most common technique is cysteine-mediated native chemical ligation (NCL). NCL requires the presence of an N-terminal cysteine on one peptide fragment and a C-terminal thioester on the other (see Ref. 46 for a summary of popular recombinant and synthetic methods for the generation of peptides bearing a C-terminal thioester) and results in the ligation of the two segments via a native peptide bond. SPPS of thioester-containing peptides has traditionally been carried out via Boc chemistry, but recent advances have enabled the robust synthesis of thioesters using the easier and more popular Fmoc chemistry[47] and commercially available Dawson Dbz resin (Novabiochem). Other means of accessing peptide thioesters via Fmoc chemistry have been recently reviewed.[48]

By strategically utilizing masked N-terminal cysteines (e.g., thioproline), multiple peptide fragments can be joined together sequentially or in a single-pot reaction.[49–53] This strategy has been used in the D-peptide synthesis of the 81-residue snow flea antifreeze protein.[54] NCL leaves a Cys residue at each ligation site, but this "scar" can be removed by desulphurization of the cysteine residue to alanine.[55,56] Furthermore, several creative adaptions of NCL allow residues other than an N-terminal cysteine to be present at the liga-

tion junction, such as N-terminal, thiol-containing auxiliary groups that can be removed via reduction,[57] UV irradiation,[58,59] or treatment with acid[60,61] after they have facilitated peptide bond formation. In another approach, modified versions of phenylalanine,[62] valine,[63] or lysine[64] bearing a thiol substituent were incorporated at the N-terminus of a peptide fragment and yielded the respective native amino acid at the ligation site following NCL/desulfurization.

Once a D-target has been synthesized, it can be used in conjunction with mirror-image display to screen peptide libraries for novel sequences of interest (see our work on HIV below and Ref. 65). The unifying feature that underlies all of the library-based display techniques discussed here is the physical linkage of a peptide to its corresponding genotype (RNA or DNA). This linkage allows the library to be subjected to multiple rounds of interrogation/library amplification leading to enrichment of sequences that bind to a target of interest. In these techniques, library diversity is generated in the nucleotide coding sequence, and cellular machinery efficiently translates this information into a peptide library. The display techniques most suitable for screening high-diversity libraries can be broken down into two broad categories: viral display and cell-free display systems (briefly described here, but for a more extensive review see Refs. 66–70).

Viral Display

Phage display continues to be the workhorse of the viral display techniques because of its ease of use, versatility, and low cost. Since phage display requires a bacterial transformation step, library size is typically limited to $\sim 10^9$–10^{10}. The most commonly utilized phage display system is the nonlytic M13-family filamentous phage, in which the peptide library is expressed as an N-terminal fusion with the pIII minor coat protein. Up to five copies of pIII are present on the phage surface, making both polyvalent and monovalent display techniques possible. Polyvalent display provides a strong avidity effect, which is highly advantageous for screening naïve peptide libraries containing only rare low affinity binders. In contrast, monovalent display reduces avidity and allows for more stringent selection of peptides with high affinity. In an early round of phage display, library diversity is high, but each sequence is represented by only a few phage. As with any library display method, the application of selection pressure must be sufficient to drive selection for tighter binders, but not so severe as to eliminate rare tight binding sequences due to stochastic factors. In later rounds, as phage library diversity drops and each remaining sequence is represented by numerous phage, selection pressure can be steadily increased. Insufficient selection pressure can select for "cheater" phage that do not bear authentic tight binding peptides (e.g., phage with growth advantages).

Besides filamentous phage display, techniques employing various eukaryotic viruses, including retroviruses, baculovirus, Adeno-associated virus, and Adenovirus have been or are currently being developed for displaying peptide libraries.[67] Other display techniques (e.g., bacterial, yeast, or mammalian cell display) have several advantages over phage display (e.g., more sophisticated folding machinery, post-translational modifications, ability to use FACS sorting), but are more complex and typically limited to less diverse libraries (reviewed by Refs. 66, 67, 71, 72).

Cell-Free Display

One of the major advantages of cell-free techniques (reviewed by Ref. 73) is that they are carried out *in vitro*. Because a transformation step is not required, library diversities $>10^{12}$ can be generated.[69] Due to the proposed correlation between library diversity and the affinity of selected ligands, this large increase in library diversity over typical viral or cell surface display systems provides a distinct advantage.

Ribosome display[74,75] capitalizes on the fact that it is possible to stall the *in vitro* translation of a polypeptide so that the ribosome remains assembled and attached to the mRNA transcript and the nascent translated polypeptide. This mRNA-ribosome-polypeptide ternary complex serves to link genotype to phenotype and can be panned against a target to isolate sequences of interest. The ternary complex can then be eluted and dissociated with EDTA, allowing for the isolation of the original mRNA transcript.

Alternatively, RNA display[76] links phenotype to genotype by connecting an mRNA sequence directly to the peptide it encodes. This linkage is accomplished by chemically attaching the antibiotic puromycin to the 3' end of the RNA via a DNA linker. As the mRNA is being translated, the ribosome will stall once it reaches the DNA linker, allowing puromycin to enter the ribosomal A site, where the ribosome catalyzes covalent attachment to the recently translated polypeptide. This peptide–RNA complex can then be subjected to panning against a specific target.

While *in vitro* display techniques that link the peptide phenotype to an RNA genotype overcome many of the limitations of phage display, the instability of RNA molecules along with other technical challenges fundamental to these techniques has limited their application to a relatively small number of expert laboratories. To address these challenges, techniques that link the library peptides directly to their encoding DNA have recently been developed.

CIS display (Isogenica) exploits the unique activity of RepA, a bacterial plasmid DNA-replication initiation protein.[77] RepA is a cis-acting protein that tightly binds to the origin of replication (*ori*) on the plasmid from which it was expressed. A stretch of DNA between the sequence that encodes RepA and the *ori* known as the CIS element contains a rho-dependent transcriptional terminator that is thought to stall the RNA polymerase during transcription of RepA. The current model holds that this delay allows the newly synthesized RepA protein emerging from the ribosome to interact with the CIS element, which subsequently directs RepA to the *ori* DNA. Peptide libraries can be fused to the N-terminus of RepA, thereby creating a link between phenotype and the DNA genotype. Like other *in vitro* techniques, CIS display has the capability to accommodate peptide libraries much larger than those possible for phage display. In one example,[77] a library of $>10^{12}$ randomized 18-mer peptides was constructed and was used to isolate sequences that bound to disparate targets. In a similar approach, DNA sequences encoding randomized peptide libraries are fused to the bacteriophage P2A gene. P2A is an endonuclease involved in the rolling circle replication of bacteriophage P2 DNA. P2A becomes covalently attached to the same DNA molecule from which it was expressed, linking phenotype to genotype. This technique has been used in a pilot study to select single-chain antibodies from a 10^7-member library and may be suitable for screening much larger libraries.[78]

D-PEPTIDE INHIBITORS OF HIV ENTRY

Here we describe the history of our potent D-peptide inhibitors of HIV entry, developed in the Kim and Kay laboratories. Initially, mirror-image polyvalent phage display was used to screen naïve peptide libraries of various lengths and geometries for binding to an HIV N-trimer pocket mimic (IQN17).[18] Pocket-specific binding was only observed in disulfide-constrained 10-mer sequences ($CX_{10}C$) containing an EWXWL consensus sequence. An initial group of ~10 winning sequences were validated by measuring their binding to the desired target and several negative control targets (mutated or missing pockets) to demonstrate pocket-specific binding.

Validated D-peptides inhibited HIV entry (lab strain HXB2) with IC_{50} values ranging from 11 to 270 μM.[18] A co-crystal structure of one of the higher affinity D-peptides (D10-p1) in complex with IQN17 shows that D10-p1 contains two short left handed α-helical segments flanking a turn imposed by the disulfide constraint. The binding interface between the hydrophobic pocket of IQN17 and D10-p1 is mediated by residues in the C-terminal α-helix, with residues in the EWXWL consensus motif making the largest contributions. Comparison of the D10-p1/IQN17 crystal structure to the native post-fusion gp41 structure[17] reveals that critical residues for binding in D10-p1 are very similar in chemical character to those of the natural C-peptide ligand (primarily W628, W631, and I635), but adopt distinct conformations due to their opposite chirality.

Due to library diversity limitations, the first-generation library only surveyed about one in a million possible sequences.[18] The identification of a strong EWXWL consensus sequence allowed us to fix these four residues to produce a "constrained" library with only six randomized residues (~10^9 possible sequences). Panning this library produced ~4-fold more potent inhibitors.[79]

Surprisingly, an 8-mer (CX_8C) was also among the winning sequences. Since 8-mers were not part of the library design and likely arose from rare replication errors, their relative success suggested that the 8-mer geometry might provide a better pocket-binding solution. Our crystal structure of the first identified 8-mer, PIE1 (pocket-specific inhibitor of entry), bound to IQN17 reveals that the key pocket-binding residues (WXWL) adopt nearly identical positions within the pocket as seen with D10-p1, leading to very similar binding interfaces despite PIE1's reduced length.[79] The key difference between PIE1 and D10-p1 is a more compact D-peptide structure with a tighter hydrophobic core devoid of water. PIE1 has a D-Pro at position 8 that likely aids making the tighter turn necessary for circularization forced by the shorter disulfide-constrained loop.[79]

To completely explore 8-mer sequence space, a new library was generated with the core consensus sequence WXWL fixed (CX_4WXWLC). While screening this library using traditional solid-phase phage display, we observed that polyvalency made it difficult to distinguish modest (μM) and tight (nM) binders. Solid-phase target presentation is advantageous for selection of weak initial binders from a naïve library, but problematic for identifying strong binders in a sea of modest binders since all binders are strongly retained on the high-density target surface. Moving the binding reaction into solution (solution-phase phage display) reduces inter-target avidity and allows additional selection pressure by reducing target concentration through rounds of panning.[80] Despite reduced inter-molecular avidity, solution-phase phage were still found to have dramatically higher binding affinities in the context of the panning than expected based on K_D values of the derived D-peptides, likely due to intra-molecular avidity on the trimeric target. To overcome this barrier, an L-peptide version of PIE2[79] (identified during earlier rounds of solution-phase phage display) was employed as a soluble competitor for subsequent rounds of

panning. Increased selection pressure was applied by escalating PIE2 concentrations, leading to the discovery of PIE7, which is ~15-fold more potent than D10-p1 (IC_{50} = 620 nM, HXB2 strain).

Our co-crystal structure of PIE7 in complex with IQN17 suggested that further gains in binding affinity could be made through optimization of the residues outside the disulfide bond, which make significant gp41 contacts.[79] Initially, these four "flanking" residues outside the disulfide bond (Gly–Ala on the N-terminus and Ala–Ala on the C-terminus) were not varied due to library cloning restrictions. We redesigned the phage display vector to relocate the cloning sites and allow randomization of the flank residues. After four rounds, PIE12 (HP-[PIE7 core]-EL) was identified with ~20-fold improved potency over PIE7. The PIE12/IQN17 crystal structure (Figure 2) reveals that PIE12's improved binding is likely due to ring-stacking interactions of D-His1 and D-Pro2 with the pocket residue Trp571 and burial of an additional 50 Å^2 hydrophobic of surface area by D-Leu15.[81] Beyond the changes in the flanking regions, the central core structure is unchanged from PIE7.

Crosslinking and the Resistance Capacitor

After battling the confounding effects of avidity throughout our phage display screens, we hoped to re-introduce avidity to boost the potency of our D-peptides. Our D-peptide/N-trimer crystal structures reveal the precise relationship between neighboring D-peptides binding to the three symmetry-related pockets. Using this information, we used discrete polyethylene glycol (PEG) crosslinkers to generate dimeric and trimeric D-peptides,[79] which showed dramatically improved antiviral potency (up to 2000-fold) over monomeric D-peptides.[79,81] PIE12-trimer, our lead inhibitor, is ~30-fold more potent than Fuzeon and inhibits a diverse panel of the most common circulating HIV strain subtypes worldwide in the high pM—low n*M* range.[81]

Interestingly, we encountered a limit to the potency gains that could be achieved by monomer affinity optimization and crosslinking. We hypothesized that this potency plateau was imposed by the limited time window available for inhibitor binding (target is only available in the transient pre-hairpin intermediate) and the inhibitor association rate (limited by diffusion), as previously observed for the pre-hairpin intermediate inhibitor 5-helix.[82] Although this potency limit would prevent us from designing more potent inhibitors, we hypothesized that "over-engineering" our inhibitors (i.e., continuing to improve inhibitor binding despite a lack of corresponding improvement in potency) would endow them with a reserve of binding energy that would stall the development of resistance mutations. We predict that this "resistance capacitor" would also greatly delay the emergence of resistance by eliminating the selective advantage conferred by these mutations (i.e., severing the link between affinity and potency). Only a profoundly disruptive mutation would escape the resistance capacitor. In support of this hypothesis, we were only able to identify high-level PIE12-trimer resistance after 65 weeks of viral passaging in the presence of inhibitor, compared with ~3 weeks for Fuzeon.[81] As predicted, PIE12-trimer was also able to absorb the impact of earlier-generation D-peptide resistance mutations.

PROTEASE-RESISTANT PEPTIDES FACE OTHER PHARMACOKINETIC CHALLENGES

Reduction of peptide susceptibility to proteases increases peptide longevity, but another major threat to serum half-life is rapid clearance via renal filtration. For globular proteins, the glomerular filtration size limit is ~70 kDa. Although albumin is slightly smaller, it avoids filtration because of electrostatic repulsion from the highly negatively charged glomerular basement membrane. Albumin is the smallest major unfiltered protein, efficiently circulating in the bloodstream with a half-life of approximately 19 days in humans.[83] The small size of peptide therapeutics means that an additional level of design is required to reduce renal filtration and realize the full benefits of protease resistance. Several common PK optimization strategies suitable for peptides are briefly described below.

PEG is a hydrophilic polymer commonly used for protein conjugation. Adding PEG to a protein has been one of the most clinically successful strategies for improving pharmacokinetics.[84] Early studies on the effects of PEG size on biodistribution revealed that good serum retention is achieved between 40 and 60 kDa, while exceeding this range resulted in increased uptake and accumulation within the reticuloendothelial system.[85] Thus, the PEGylation field has largely adopted the strategy of adding ~40 kDa of PEG weight to peptide and small protein therapeutics. PEG is extensively hydrated such that its hydrodynamic radius is much larger than expected from its molecular weight. Furthermore, distributing the weight of the PEG polymer in a branched geometry improves half-life and reduces steric interference.[86] PEG conjugation can also be reversible (e.g., an ester linkage), creating a circulating depot from which the therapeutic is cleaved over time (e.g., in case drug activity is adversely affected by PEG conjugation).[87,88] Limitations of PEGylation include steric interference with binding, long-lived accumulation in renal tubule cells, viscosity, and polydispersity. An alternative approach uses a hydroxyethyl starch polymer (HESylation) to reduce renal filtration.[89]

Albumin binding (covalent or noncovalent) is another recently validated approach for prolonging serum half-life (reviewed by Ref. 90). Promising albumin-binding strategies include covalent albumin-peptide conjugation, as well as reversible binding to circulating albumin via albumin-binding peptides, small molecules, or fatty acids.[90–92] As an example, albumin conjugation of an HIV C-peptide inhibitor (either *in vitro* or *in vivo*) dramatically improves serum half-life,[93] as does cholesterol conjugation to a lesser extent, presumably via weak reversible interactions with albumin and/or cell membranes.[94]

FUTURE DIRECTIONS

Recombinant Production of Peptidomimetics

Although robust recombinant production of peptidomimetics is not yet possible, significant recent advances in synthetic biology may enable routine production of diverse peptidomimetic libraries in the near future. One promising approach is *in vitro* codon reprogramming for the synthesis of unnatural polymers. This approach relies on cell-free translation systems to reconstitute ribosomal peptide synthesis using a minimal set of purified protein components.[95–100] By chemically or enzymatically charging tRNA molecules with novel amino acid analogues, the genetic code can be effectively reprogrammed *in vitro*. When these cell-free systems with genetic code modifications are used in conjunction with a display technology, peptides with novel amino acids can be screened for a desired property. For example, ribosome display was used in conjunction with *in vitro* codon reprogramming to isolate peptide sequences from an mRNA library that encoded an unnatural, selectable amino acid.[101–103]

Along these lines, it has been demonstrated that tRNAs can be charged with a variety of amino acid analogues that will modify the peptide backbone, including α-hydroxy acids, N-methyl amino acids, α,α-disubstituted amino acids, β-amino acids, and D-amino acids.[104] However, the efficiency of ribosomal incorporation of Ala/Phe analogues varies greatly from fairly robust (α-hydroxy acid and N-methyl) to weak (α,α-disubstituted amino acids) to undetectable (β- and D-amino acids).[104] Subsequent work has described the ability of the translation machinery to accommodate amino acid analogues with novel side chains and backbones.[105]

In one example, seven codons were each reassigned to encode a unique α-hydroxy acid, and polymers as long as 12 consecutive α-hydroxy acids could be synthesized.[106] In another report, the incorporation efficiencies of 23 N-methyl amino acids, 19 of which bore naturally occurring side chains, were determined. Eight of these 19 N-methyl amino acids were incorporated at specific points in a polypeptide with >30% efficiency as compared with wild type. A peptide up to 10 residues long could be synthesized from three unique N-methyl amino acids.[107]

While less success has been reported with ribosomal incorporation of D-amino acids, modifications to the ribosomal peptidyltransferase center and helix 89 of the 23S rRNA can relax the ribosome's natural substrate specificity, thereby enhancing the incorporation of D-amino acid residues into a growing polypeptide chain.[108,109] Although these techniques have not yet been employed as such, in principle cell-free translation systems coupled with *in vitro* display techniques could be used to screen libraries of polymers with novel backbones. As an advance in this direction, genetic code reprogramming has already been used in conjunction with mRNA display technology to generate mRNA-peptide fusions containing N-methyl amino acids.[110]

Another approach to recombinantly produce peptidomimetics relies on the ability to expand the genetic code *in vivo* via the generation of evolved tRNA/aminoacyl-tRNA synthetase pairs. In these systems, the foreign tRNA functions as an amber suppressor, effectively allowing the amber nonsense codon to be reprogrammed to encode a non-natural amino acid.[111,112] It has been demonstrated that genetic code expansion can be used in conjunction with phage display to incorporate a non-natural amino acid into a pIII fusion peptide.[113] In the future, multiple codons could be reassigned, permitting the incorporation of multiple unnatural residues *in vivo*. Several advances have been made toward this end. In a recent publication describing a technique for rapid, genome-wide engineering, the authors show progress toward replacing all 314 TAG stop codons in *E. coli* with the TAA stop codon.[114] This type of genome manipulation could be used for the removal of redundancy from the genetic code, freeing up codons for potential reprogramming. In another approach involving evolved tRNA/aminoacyl-tRNA synthetase pairs, an evolved orthogonal ribosome able to read both 3- and 4-base codons was able to efficiently incorporate two different non-natural amino acids into a single polypeptide chain *in vivo*.[115,116] These approaches present tantalizing possibilities for the production of peptide libraries with unnatural side chains and backbones, but the technology is not yet sufficiently robust to allow for widespread application. Additional engineering of tRNA molecules, elongation factors, and the ribosome itself will likely be required for use with certain diverse peptidomimetics.[109,117,118]

D-Peptides present a unique opportunity for designing an artificial recombinant production system. Because of their symmetry relationship with natural peptides, an *in vitro*

translation system composed of all opposite-chirality components (D-proteins and nucleotides containing L-ribose) would function equivalently to natural translation, when provided with mirror-image DNA substrates. Synthesis of all ribosomal components presents an enormous synthetic challenge, but recent advances in SPPS and peptide ligation may now make this approach feasible. A mirror-image *in vitro* translation system would provide a useful tool for D-peptide drug discovery and production, but may not be ideal for large-scale production, especially of complex D-proteins (e.g., those requiring chaperones or post-translational modifications). The ultimate goal is to produce D-peptides using a synthetic mirror-image organism, a strategy we dub the "D. coli" project. The key to this project is synthesizing the minimal set of RNAs and proteins necessary to allow enzymatic production of other larger components and ultimately all components needed for a self-replicating organism. It is also not yet clear how to "start up" such an organism.[119,120]

Cost and Toxicity of Peptidomimetics

In addition to achieving their biological objectives, peptidomimetics will need to overcome concerns about cost and toxicity to succeed as therapeutics. Currently there are no FDA-approved fully peptidomimetic peptides, so information on their *in vivo* toxicity is extremely limited. Initial data from two D-peptides that have advanced to clinical trials (Genzyme's Delmitide[121] and Allelix's ALX40-4C[122]) showed that both D-peptides (one orally administered, one systemically delivered) were well tolerated in humans. Further comfort is provided by over a dozen approved D-amino acid-containing peptides, as well as two approved ß-amino acid-containing peptides.[123] These data suggest that these amino acids are not intrinsically toxic, but more rigorous animal toxicology studies on different classes of fully protease-resistant peptides will be required for a definitive determination. Such studies will also determine whether these peptidomimetics induce significant immunogenicity upon chronic administration. Finally, the cost of D-, *β*-, and other uncommon amino acids is currently significantly higher than the corresponding common L-amino acids, largely because of their current status as specialty reagents. However, we anticipate the cost of these amino acids will drop dramatically as they are adopted in high-volume production of therapeutic peptides, as has already occurred with several D-amino acids in large-scale peptide production.

The authors thank Debra Eckert for critical review of the article and figures preparation. M.S.K. is a Scientific Director and consultant of the D-peptide Research Division of Navigen, which is commercializing D-peptide inhibitors of viral entry.

REFERENCES

1. McGregor, D. P. Curr Opin Pharmacol 2008, 8, 616–619.
2. Patch, J. A.; Barron, A. E. Curr Opin Chem Biol 2002, 6, 872–877.
3. Goodman, C. M.; Choi, S.; Shandler, S.; DeGrado, W. F. Nat Chem Biol 2007, 3, 252–262.
4. Hill, D. J.; Mio, M. J.; Prince, R. B.; Hughes, T. S.; Moore, J. S. Chem Rev 2001, 101, 3893–4012.
5. Stevenson, C. L. Curr Pharm Biotechnol 2009, 10, 122–137.
6. Tilton, J. C.; Doms, R. W. Antiviral Res 2010, 85, 91–100.
7. Eckert, D. M.; Kim, P. S. Annu Rev Biochem 2001, 70, 777–810.
8. Chan, D. C.; Kim, P. S. Cell 1998, 93, 681–684.
9. Jiang, S.; Lin, K.; Strick, N.; Neurath, A. R. Nature 1993, 365, 113.
10. Wild, C.; Oas, T.; McDanal, C.; Bolognesi, D.; Matthews, T. Proc Natl Acad Sci U S A 1992, 89, 10537–10541.
11. Wild, C. T.; Shugars, D. C.; Greenwell, T. K.; McDanal, C. B.; Matthews, T. J. Proc Natl Acad Sci U S A 1994, 91, 9770–9774.
12. Lu, M.; Blacklow, S. C.; Kim, P. S. Nat Struct Biol 1995, 2, 1075–1082.
13. Eckert, D. M.; Kim, P. S. Proc Natl Acad Sci U S A 2001, 98, 11187–11192.
14. Root, M. J.; Kay, M. S.; Kim, P. S. Science 2001, 291, 884–888.
15. Kilby, J. M.; Hopkins, S.; Venetta, T. M.; DiMassimo, B.; Cloud, G. A.; Lee, J. Y.; Alldredge, L.; Hunter, E.; Lambert, D.; Bolognesi, D.; Matthews, T.; Johnson, M. R.; Nowak, M. A.; Shaw, G. M.; Saag, M. S. Nat Med 1998, 4, 1302–1307.
16. Chan, D. C.; Chutkowski, C. T.; Kim, P. S. Proc Natl Acad Sci U S A 1998, 95, 15613–15617.
17. Chan, D. C.; Fass, D.; Berger, J. M.; Kim, P. S. Cell 1997, 89, 263–273.
18. Eckert, D. M.; Malashkevich, V. N.; Hong, L. H.; Carr, P. A.; Kim, P. S. Cell 1999, 99, 103–115.
19. Frey, G.; Rits-Volloch, S.; Zhang, X. Q.; Schooley, R. T.; Chen, B.; Harrison, S. C. Proc Natl Acad Sci U S A 2006, 103, 13938–13943.
20. Ferrer, M.; Kapoor, T. M.; Strassmaier, T.; Weissenhorn, W.; Skehel, J. J.; Oprian, D.; Schreiber, S. L.; Wiley, D. C.; Harrison, S. C. Nat Struct Biol 1999, 6, 953–960.
21. Debnath, A. K.; Radigan, L.; Jiang, S. J Med Chem 1999, 42, 3203–3209.
22. Jin, B. S.; Ryu, J. R.; Ahn, K.; Yu, Y. G. AIDS Res Hum Retroviruses 2000, 16, 1797–1804.
23. Jiang, S.; Lu, H.; Liu, S.; Zhao, Q.; He, Y.; Debnath, A. K. Antimicrob Agents Chemother 2004, 48, 4349–4359.
24. Dwyer, J. J.; Wilson, K. L.; Davison, D. K.; Freel, S. A.; Seedorff, J. E.; Wring, S. A.; Tvermoes, N. A.; Matthews, T. J.; Greenberg, M. L.; Delmedico, M. K. Proc Natl Acad Sci U S A 2007, 104, 12772–12777.
25. Lalezari, J. P.; Bellos, N. C.; Sathasivam, K.; Richmond, G. J.; Cohen, C. J.; Myers, R. A.; , Jr.Henry, D. H.; Raskino, C.; Melby, T.; Murchison, H.; Zhang, Y.; Spence, R.; Greenberg, M. L.; Demasi, R. A.; Miralles, G. D. J Infect Dis 2005, 191, 1155–1163.
26. Ray, N.; Harrison, J. E.; Blackburn, L. A.; Martin, J. N.; Deeks, S. G.; Doms, R. W. J Virol 2007, 81, 3240–3250.
27. Cheng, R. P.; Gellman, S. H.; DeGrado, W. F. Chem Rev 2001, 101, 3219–3232.

28. Pilsl, L. K.; Reiser, O. Amino acids 2011, 41, 709–718.
29. Horne, W. S.; Gellman, S. H. Acc Chem Res 2008, 41, 1399–1408.
30. Seebach, D.; Gardiner, J. Acc Chem Res 2008, 41, 1366–1375.
31. Kritzer, J. A.; Stephens, O. M.; Guarracino, D. A.; Reznik, S. K.; Schepartz, A. Biorg Med Chem 2005, 13, 11–16.
32. Harker, E. A.; Daniels, D. S.; Guarracino, D. A.; Schepartz, A. Biorg Med Chem 2009, 17, 2038–2046.
33. Stephens, O. M.; Kim, S.; Welch, B. D.; Hodsdon, M. E.; Kay, M. S.; Schepartz, A. J Am Chem Soc 2005, 127, 13126–13127.
34. Bautista, A. D.; Stephens, O. M.; Wang, L.; Domaoal, R. A.; Anderson, K. S.; Schepartz, A. Bioorg Med Chem Lett 2009, 19, 3736–3738.
35. English, E. P.; Chumanov, R. S.; Gellman, S. H.; Compton, T. J Biol Chem 2006, 281, 2661–2667.
36. Horne, W. S.; Johnson, L. M.; Ketas, T. J.; Klasse, P. J.; Lu, M.; Moore, J. P.; Gellman, S. H. Proc Natl Acad Sci U S A 2009, 106, 14751–14756.
37. Horne, W. S.; Price, J. L.; Keck, J. L.; Gellman, S. H. J Am Chem Soc 2007, 129, 4178–4180.
38. Johnson, L. M.; Horne, W. S.; Gellman, S. H. J Am Chem Soc 2011, 133, 10038–10041.
39. Lam, K. S.; Lehman, A. L.; Song, A.; Doan, N.; Enstrom, A. M.; Maxwell, J.; Liu, R. Methods Enzymol 2003, 369, 298–322.
40. Zawadzke, L. E.; Berg, J. M. J Am Chem Soc 1992, 114, 4002–4003.
41. Milton, R. C.; Milton, S. C.; Kent, S. B. Science 1992, 256, 1445–1448.
42. Schumacher, T. N.; Mayr, L. M.; Minor, D. L.; Jr. Milhollen, M. A.; Burgess, M. W.; Kim, P. S. Science 1996, 271, 1854–1857.
43. Kent, S. B. Chem Soc Rev 2009, 38, 338–351.
44. Clark-Lewis, I.; Aebersold, R.; Ziltener, H.; Schrader, J. W.; Hood, L. E.; Kent, S. B. Science 1986, 231, 134–139.
45. Hackenberger, C. P.; Schwarzer, D. Angew Chem Int Ed 2008, 47, 10030–10074.
46. Muralidharan, V.; Muir, T. W. Nat Methods 2006, 3, 429–438.
47. Blanco-Canosa, J. B.; Dawson, P. E. Angew Chem Int Ed 2008, 47, 6851–6855.
48. Mende, F.; Seitz, O. Angew Chem Int Ed 2011, 50, 1232–1240.
49. Bang, D.; Chopra, N.; Kent, S. B. H. J Am Chem Soc 2004, 126, 1377–1383.
50. Bang, D.; Kent, S. B. H. Agnew Chem Int Ed 2004, 43, 2534–2538.
51. Boerema, D. J.; Tereshko, V. A.; Kent, S. B. H. Biopolymers 2008, 90, 278–286.
52. Mandal, K.; Kent, S. B. Angew Chem Int Ed 2011, 50, 8029–8033.
53. Kumar, K. S.; Bavikar, S. N.; Spasser, L.; Moyal, T.; Ohayon, S.; Brik, A. Angew Chem Int Ed 2011, 50, 6137–6141.
54. Pentelute, B. L.; Gates, Z. P.; Dashnau, J. L.; Vanderkooi, J. M.; Kent, S. B. H. J Am Chem Soc 2008, 130, 9702–9707.
55. Wan, Q.; Danishefsky, S. J. Angew Chem 2007, 46, 9248–9252.
56. Rohde, H.; Seitz, O. Biopolymers 2010, 94, 551–559.
57. Canne, L. E.; Bark, S. J.; Kent, S. B. H. J Am Chem Soc 1996, 118, 5891–5896.
58. Kawakami, T.; Aimoto, S. Tetrahedron Lett 2003, 44, 6059–6061.
59. Marinzi, C.; Offer, J.; Longhi, R.; Dawson, P. E. Biorg Med Chem 2004, 12, 2749–2757.
60. Botti, P.; Carrasco, M. R.; Kent, S. B. H. Tetrahedron Lett 2001, 42, 1831–1833.
61. Offer, J.; Boddy, C. N.; Dawson, P. E. J Am Chem Soc 2002, 124, 4642–4646.
62. Crich, D.; Banerjee, A. J Am Chem Soc 2007, 129, 10064–10065.
63. Haase, C.; Rohde, H.; Seitz, O. Angew Chem 2008, 47, 6807–6810.
64. Yang, R.; Pasunooti, K. K.; Li, F.; Liu, X. W.; Liu, C. F. Chem Commun (Camb) 2010, 46, 7199–7201.
65. Liu, M.; Pazgier, M.; Li, C.; Yuan, W.; Lu, W. Angew Chem Int Ed 2010, 49, 3649–3652.
66. Gronwall, C.; Stahl, S. J Biotechnol 2009, 140, 254–269.
67. Sergeeva, A.; Kolonin, M. G.; Molldrem, J. J.; Pasqualini, R.; Arap, W. Adv Drug Deliv Rev 2006, 58, 1622–1654.
68. Kehoe, J. W.; Kay, B. K. Chem Rev 2005, 105, 4056–4072.
69. FitzGerald, K. Drug Discovery Today 2000, 5, 253–258.
70. Ullman, C. G.; Frigotto, L.; Cooley, R. N. Brief Funct Genomics 2011, 10, 125–134.
71. Lee, S. Y.; Choi, J. H.; Xu, Z. H. Trends Biotechnol 2003, 21, 45–52.
72. Daugherty, P. S. Curr Opin Struct Biol 2007, 17, 474–480.
73. Barendt, P. A.; Sarkar, C. A. In Protein Engineering and Design; Park, S. F.; Cochran, J. R., Eds.; CRC Press, 2009; pp 51–82.
74. Mattheakis, L. C.; Bhatt, R. R.; Dower, W. J. Proc Natl Acad Sci U S A 1994, 91, 9022–9026.
75. Hanes, J.; Pluckthun, A. Proc Natl Acad Sci U S A 1997, 94, 4937–4942.
76. Roberts, R. W.; Szostak, J. W. Proc Natl Acad Sci U S A 1997, 94, 12297–12302.
77. Odegrip, R.; Coomber, D.; Eldridge, B.; Hederer, R.; Kuhlman, P. A.; Ullman, C.; FitzGerald, K.; McGregor, D. Proc Natl Acad Sci U S A 2004, 101, 2806–2810.
78. Reiersen, H.; Lobersli, I.; Loset, G. A.; Hvattum, E.; Simonsen, B.; Stacy, J. E.; McGregor, D.; FitzGerald, K.; Welschof, M.; Brekke, O. H.; Marvik, O. J. Nucleic Acids Res 2005, 33, e10.
79. Welch, B. D.; VanDemark, A. P.; Heroux, A.; Hill, C. P.; Kay, M. S. Proc Natl Acad Sci U S A 2007, 104, 16828–16833.
80. Barbas, C. F. Phage Display: A Laboratory Manual; Cold Springs Harbor Laboratory Press: New York, 2001.
81. Welch, B. D.; Francis, J. N.; Redman, J. S.; Paul, S.; Weinstock, M. T.; Reeves, J. D.; Lie, Y. S.; Whitby, F. G.; Eckert, D. M.; Hill, C. P.; Root, M. J.; Kay, M. S. J Virol 2010, 84, 11235–11244.
82. Steger, H. K.; Root, M. J. J Biol Chem 2006, 281, 25813–25821.
83. Dennis, M. S.; Zhang, M.; Meng, Y. G.; Kadkhodayan, M.; Kirchhofer, D.; Combs, D.; Damico, L. A. J Biol Chem 2002, 277, 35035–35043.
84. Fishburn, C. S. J Pharm Sci 2008, 97, 4167–4183.
85. Yamaoka, T.; Tabata, Y.; Ikada, Y. J Pharm Sci 1994, 83, 601–606.
86. Fee, C. J. Biotechnol Bioeng 2007, 98, 725–731.
87. Greenwald, R. B. J Control Release 2001, 74, 159–171.
88. Tong, R.; Cheng, J. Polym Rev 2007, 47, 345–381.
89. Besheer, A.; Hertel, T. C.; Kressler, J.; Mader, K.; Pietzsch, M. Methods Mol Biol 2011, 751, 17–27.
90. Kratz, F. J Control Release 2008, 132, 171–183.

91. Madsen, K.; Knudsen, L. B.; Agersoe, H.; Nielsen, P. F.; Thogersen, H.; Wilken, M.; Johansen, N. L. J Med Chem 2007, 50, 6126–6132.
92. Trussel, S.; Dumelin, C.; Frey, K.; Villa, A.; Buller, F.; Neri, D. Bioconjug Chem 2009, 20, 2286–2292.
93. Stoddart, C. A.; Nault, G.; Galkina, S. A.; Thibaudeau, K.; Bakis, P.; Bousquet-Gagnon, N.; Robitaille, M.; Bellomo, M.; Paradis, V.; Liscourt, P.; Lobach, A.; Rivard, M. E.; Ptak, R. G.; Mankowski, M. K.; Bridon, D.; Quraishi, O. J Biol Chem 2008, 283, 34045–34052.
94. Ingallinella, P.; Bianchi, E.; Ladwa, N. A.; Wang, Y. J.; Hrin, R.; Veneziano, M.; Bonelli, F.; Ketas, T. J.; Moore, J. P.; Miller, M. D.; Pessi, A. Proc Natl Acad Sci U S A 2009, 106, 5801–5806.
95. Forster, A. C.; Weissbach, H.; Blacklow, S. C. Anal Biochem 2001, 297, 60–70.
96. Forster, A. C.; Tan, Z. P.; Nalam, M. N. L.; Lin, H. N.; Qu, H.; Cornish, V. W.; Blacklow, S. C. Proc Natl Acad Sci U S A 2003, 100, 6353–6357.
97. Shimizu, Y.; Kanamori, T.; Ueda, T. Methods 2005, 36, 299–304.
98. Shimizu, Y.; Kuruma, Y.; Ying, B. W.; Umekage, S.; Ueda, T. FEBS J 2006, 273, 4133–4140.
99. Ohta, A.; Yamagishi, Y.; Suga, H. Curr Opin Chem Biol 2008, 12, 159–167.
100. Shimizu, Y.; Inoue, A.; Tomari, Y.; Suzuki, T.; Yokogawa, T.; Nishikawa, K.; Ueda, T. Nat Biotechnol 2001, 19, 751–755.
101. Tan, Z. P.; Blacklow, S. C.; Cornish, V. W.; Forster, A. C. Methods 2005, 36, 279–290.
102. Forster, A. C.; Cornish, V. W.; Blacklow, S. C. Anal Biochem 2004, 333, 358–364.
103. Watts, R. E.; Forster, A. C. Methods Mol Biol 2012, 805, 349–365.
104. Tan, Z. P.; Forster, A. C.; Blacklow, S. C.; Cornish, V. W. J Am Chem Soc 2004, 126, 12752–12753.
105. Hartman, M. C.; Josephson, K.; Lin, C. W.; Szostak, J. W. PloS One 2007, 2, e972.
106. Ohta, A.; Murakami, H.; Suga, H. Chembiochem 2008, 9, 2773–2778.
107. Kawakami, T.; Murakami, H.; Suga, H. Chem Biol 2008, 15, 32–42.
108. Dedkova, L. M.; Fahmi, N. E.; Golovine, S. Y.; Hecht, S. M. J Am Chem Soc 2003, 125, 6616–6617.
109. Dedkova, L. M.; Fahmi, N. E.; Golovine, S. Y.; Hecht, S. M. Biochemistry 2006, 45, 15541–15551.
110. Frankel, A.; Millward, S. W.; Roberts, R. W. Chem Biol 2003, 10, 1043–1050.
111. Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G. Science 2001, 292, 498–500.
112. Wang, Q.; Parrish, A. R.; Wang, L. Chem Biol 2009, 16, 323–336.
113. Feng, T.; Tsao, M. L.; Schultz, P. G. J Am Chem Soc 2004, 126, 15962–15963.
114. Isaacs, F. J.; Carr, P. A.; Wang, H. H.; Lajoie, M. J.; Sterling, B.; Kraal, L.; Tolonen, A. C.; Gianoulis, T. A.; Goodman, D. B.; Reppas, N. B.; Emig, C. J.; Bang, D.; Hwang, S. J.; Jewett, M. C.; Jacobson, J. M.; Church, G. M. Science 2011, 333, 348–353.
115. Neumann, H.; Wang, K.; Davis, L.; Garcia-Alai, M.; Chin, J. W. Nature 2010, 464, 441–444.
116. Wang, K.; Schmied, W. H.; Chin, J. W. Angew Chem Int Ed 2012, 51, 2288–2297.
117. Doi, Y.; Ohtsuki, T.; Shimizu, Y.; Ueda, T.; Sisido, M. J Am Chem Soc 2007, 129, 14458–14462.
118. Dale, T.; Uhlenbeck, O. C. Trends Biochem Sci 2005, 30, 659–665.
119. Jewett, M. C.; Forster, A. C. Curr Opin Biotechnol 2010, 21, 697–703.
120. Forster, A. C.; Church, G. M. Mol Syst Biol 2006, 2, 45.
121. Travis, S.; Yap, L. M.; Hawkey, C.; Warren, B.; Lazarov, M.; Fong, T.; Tesi, R. J. Inflamm Bowel Dis 2005, 11, 713–719.
122. Doranz, B. J.; Filion, L. G.; Diaz-Mitoma, F.; Sitar, D. S.; Sahai, J.; Baribaud, F.; Orsini, M. J.; Benovic, J. L.; Cameron, W.; Doms, R. W. AIDS Res Hum Retroviruses 2001, 17, 475–486.
123. Vlieghe, P.; Lisowski, V.; Martinez, J.; Khrestchatisky, M. Drug Discov Today 2010, 15, 40–56.

CHAPTER 3

DESIGN OF A POTENT D-PEPTIDE HIV-1 ENTRY INHIBITOR WITH A STRONG BARRIER TO RESISTANCE

Brett D. Welch*, J. Nicholas Francis*, Joseph S. Redman, Suparna Paul, Matthew T. Weinstock, Jacqueline D. Reeves, Yolanda S. Lie, Frank G. Whitby, Debra M. Eckert, Christopher P. Hill, Michael J. Root, Michael S. Kay

* These authors contributed equally to this work

Reprinted with permission from The Journal of Virology,

Vol. 84, Issue 21, pp. 11235-11244

JOURNAL OF VIROLOGY, Nov. 2010, p. 11235–11244 Vol. 84, No. 21
0022-538X/10/$12.00 doi:10.1128/JVI.01339-10

Design of a Potent D-Peptide HIV-1 Entry Inhibitor with a Strong Barrier to Resistance▽

Brett D. Welch,[1]† J. Nicholas Francis,[1]† Joseph S. Redman,[1] Suparna Paul,[2] Matthew T. Weinstock,[1] Jacqueline D. Reeves,[3] Yolanda S. Lie,[3] Frank G. Whitby,[1] Debra M. Eckert,[1] Christopher P. Hill,[1] Michael J. Root,[2] and Michael S. Kay[1]*

Department of Biochemistry, University of Utah School of Medicine, Salt Lake City, Utah 84112[1]; Department of Biochemistry and Molecular Biology, Thomas Jefferson University, Philadelphia, Pennsylvania 19107[2]; and Monogram Biosciences, 345 Oyster Point Blvd., South San Francisco, California 94080[3]

Received 23 June 2010/Accepted 6 August 2010

The HIV gp41 N-trimer pocket region is an ideal viral target because it is extracellular, highly conserved, and essential for viral entry. Here, we report on the design of a pocket-specific D-peptide, PIE12-trimer, that is extraordinarily elusive to resistance and characterize its inhibitory and structural properties. D-Peptides (peptides composed of D-amino acids) are promising therapeutic agents due to their insensitivity to protease degradation. PIE12-trimer was designed using structure-guided mirror-image phage display and linker optimization and is the first D-peptide HIV entry inhibitor with the breadth and potency required for clinical use. PIE12-trimer has an ultrahigh affinity for the gp41 pocket, providing it with a reserve of binding energy (resistance capacitor) that yields a dramatically improved resistance profile compared to those of other fusion inhibitors. These results demonstrate that the gp41 pocket is an ideal drug target and establish PIE12-trimer as a leading anti-HIV antiviral candidate.

The HIV envelope protein (Env) mediates viral entry into cells (11). Env is cleaved into surface (gp120) and transmembrane (gp41) subunits that remain noncovalently associated to form trimeric spikes on the virion surface (16). gp120 recognizes target cells by interacting with cellular receptors, while gp41 mediates membrane fusion. Peptides derived from heptad repeats near the N and C termini of the gp41 ectodomain (N and C peptides) interact in solution to form a six-helix bundle, representing the postfusion structure (3, 55, 56). In this structure, N peptides form a central trimeric coiled coil (N trimer), creating grooves into which C peptides bind. This structure, in conjunction with the dominant-negative inhibitory properties of exogenous N and C peptides, suggests a mechanism for Env-mediated entry (10, 22, 58–60).

During entry, gp41 forms an extended prehairpin intermediate that leaves the exposed N-trimer region vulnerable to inhibition for several minutes (18, 35). This intermediate ultimately collapses as the C-peptide regions bind to the N-trimer grooves to form a trimer of hairpins (six-helix bundle), juxtaposing viral and cellular membranes and inducing fusion. Enfuvirtide (Fuzeon), the only clinically approved HIV fusion inhibitor, is a C peptide that binds to part of the N-trimer groove and prevents six-helix bundle formation in a dominant-negative manner (61). Enfuvirtide is active in patients with multidrug resistance to other classes of inhibitors and is a life-prolonging option for these patients (30, 31). However, enfuvirtide use is restricted to salvage therapy due to several limitations, including (i) high dosing requirements (90 mg, twice-daily injections), (ii) high cost (~$30,000/year/patient in the United States), and (iii) the rapid emergence of resistant strains (21, 47).

A deep hydrophobic pocket at the base of the N-trimer groove is an especially attractive inhibitory target because of its high degree of conservation (3, 12, 48), poor tolerance to substitution (4, 34), and critical role in membrane fusion (2). Indeed, this region is conserved at both the amino acid level (for gp41 function in membrane fusion) and the nucleotide level (for the structured RNA region of the Rev-responsive element). Enfuvirtide binds to the N-trimer groove just N terminal to the pocket and is significantly more susceptible to resistance mutations than 2nd-generation C-peptide inhibitors, such as T-1249, that also bind to the pocket (8, 13, 29, 44, 46, 47, 58).

Peptide design, molecular modeling, and small-molecule screening have produced a diverse set of compounds that interact with the gp41 pocket and inhibit HIV-1 entry with modest potency, but often with significant cytotoxicity (7, 14, 15, 17, 23, 24, 26, 34, 51, 54). The first direct evidence that pocket-specific binders are sufficient to inhibit HIV entry came with the discovery of protease-resistant D-peptides identified using mirror-image phage display (12). In this technique, a phage library is screened against a mirror-image version of the target protein (synthesized using D-amino acids) (50). By symmetry, mirror images (D-peptides) of the discovered sequences will bind to the natural L-peptide target. As the mirror images of naturally occurring L-peptides, D-peptides cannot be digested by natural proteases. Protease resistance provides D-peptides theoretical treatment advantages of extended survival in the body and possible oral bioavailability (41, 42, 49).

* Corresponding author. Mailing address: Department of Biochemistry, University of Utah School of Medicine, 15 N. Medical Drive East, Rm. 4100, Salt Lake City, UT 84112-5650. Phone: (801) 585-5021. Fax: (801) 581-7959. E-mail: kay@biochem.utah.edu.
† These authors contributed equally to this work.
▽ Published ahead of print on 18 August 2010.

These 1st-generation D-peptide entry inhibitors possess potency against a laboratory-adapted isolate (HXB2) at low to mid-μM concentrations (12). We previously reported an affinity-matured 2nd-generation D-peptide called PIE7, *p*ocket-specific *i*nhibitor of *e*ntry 7 (57). A trimeric version of PIE7 is the first high-affinity pocket-specific HIV-1 inhibitor and has potency against X4-tropic (HXB2) and R5-tropic (BaL) strains at sub-nM concentrations. However, significant further optimization is required to create a robust clinical candidate for two reasons. First, this D-peptide is much less potent (requiring high nM concentrations) against JRFL, a primary R5-tropic strain. Therefore, improved PIE potency is necessary to combat diverse primary strains. Second, by improving the affinity of our inhibitors for the pocket target, we hope to provide a reserve of binding energy that will delay the emergence of drug resistance, as described below.

We and others have reported a potency plateau for some gp41-based fusion inhibitors that is likely imposed by the transient exposure of the prehairpin intermediate (9, 27, 53, 57). For very high-affinity inhibitors, association kinetics (rather than affinity) limits potency so that two inhibitors with significantly different affinities for the prehairpin intermediate can have similar antiviral potencies. We proposed that overengineering our D-peptides with substantial affinity beyond this potency plateau would provide a reserve of binding energy that would combat affinity-disrupting resistance mutations (57). Such a resistance capacitor should also prevent the stepwise accumulation of subtle resistance mutations in Env by eliminating the selective advantage that such mutants would otherwise confer.

Here, we report on the design and characterization of a 3rd-generation pocket-specific D-peptide, PIE12-trimer, with ~100,000-fold improved target binding compared to that of the best previous D-peptide, significantly broadened inhibitory potency, and an enhanced resistance capacitor that provides a strong barrier to viral resistance. We achieved this increased potency via structure-guided phage display and crosslinker optimization. PIE12-trimer has a dramatically improved resistance profile compared to the profiles of earlier D-peptides, as well as those of enfuvirtide and T-1249. These results validate the resistance capacitor hypothesis and establish PIE12-trimer as a leading anti-HIV therapeutic candidate.

MATERIALS AND METHODS

Peptide synthesis. All peptides were synthesized as described previously (57). All dimers and trimers except PIE12-trimer were made essentially as described using bis-dPEG$_5$ NHS ester (where PEG is polyethylene glycol and NHS is *N*-hydroxysuccinimide; catalog no. 10224; Quanta BioDesign); PIE12-trimer was synthesized using the following higher-yield protocol. PIE12-GK (2 mM) was reacted with bis-dPEG$_5$ NHS ester crosslinker (1 M stock in dimethylacetamide) at a 1:20 (peptide/PEG) molar ratio in 100 mM HEPES (pH 7.8 to 8) for 90 s at room temperature (RT). The reaction was stopped by addition of acetic acid to 5% and 3 M guanidine HCl (GuHCl) and purified by reverse-phase high-pressure liquid chromatography (RP-HPLC; C$_{18}$ column; Vydac). This product (~3 to 5 mM) was reacted at a 2:1 molar excess with PIE12-GKK in dimethylacetamide buffered by triethylamine (pH 7.5) for 75 min and purified by RP-HPLC (C$_{18}$ column; Vydac).

Phage display vector design. Use of a commercially available phage library cloning system (NEB) allowed us to relocate cloning sites away from the flanking regions (38). We redesigned the regions immediately outside the flanking residues in our cloning vector in order to structurally isolate them and minimize any bias caused by flanking sequence randomization. Our library peptides are displayed as fusions to the phage p3 protein, which contains an N-terminal leader sequence that is cleaved by *Escherichia coli* secretion signal peptidases. In the original vector, the N-terminal flanking residues of the library peptides are immediately adjacent to the secretion signal. Due to proximity to the secretion signal cleavage site, it is likely that randomization of these residues would differentially affect library-p3 protein secretion and peptide presentation on the phage surface. This bias would confound the selection of N-terminal flanking sequences solely on the basis of their affinity for the N trimer. To avoid this bias, we introduced a five-amino-acid spacer to structurally isolate the cleavage site from the randomized N-terminal flanking residues. We choose the N-terminal residues (KIEEG) from maltose binding protein (MBP) as the spacer sequence, since MBP is very efficiently cleaved during secretion from *E. coli*.

We have observed that mutations in the C-terminal sequence that links the peptide to the phage p3 protein can also create undesirable selection bias (presumably by allowing the C terminus of the D-peptides to form a continuous helix with the N terminus of p3, thus enhancing peptide presentation to the target) (57). Therefore, a flexible GGGS spacer was inserted after the C-terminal flanking residues to structurally isolate them from the N terminus of p3.

To validate this new phage display vector, we used it to clone an earlier PIE (PIE2) along with a mutant (PIE2-AAA) which had previously been observed to enhance phage affinity for the pocket target via mutation of the linker between the library peptide and p3, although this mutation did not enhance inhibitor potency when incorporated into a D-peptide (57). We assayed the target binding affinity of the resultant phage (Φ) and compared it to that of phage produced with the previous phage vector. In the previous phage vector, PIE2-AAA-Φ "cheated" in order to bind to the target with an ~70-fold more affinity than PIE2-Φ, but this difference was abolished in the modified vector (data not shown). Furthermore, sequencing revealed that N-terminal flanking residues from the amplified phage library prior to selection were random, indicating a lack of bias due to signal peptidase cleavage efficiency.

Phage display. An 8-mer flanking library phage display was performed essentially as described previously (57). Four rounds of mirror-image solution-phase phage display were performed by incubating (for 2 h at RT) 10^{10} phage (amplified from the previous round) with 10 nM biotinylated D-IZN17 (a mimic of the D-peptide gp41 pocket target) in the presence of escalating soluble competitor (L-2K-PIE2) (10, 30, 90, and 360 μM for rounds 1 to 4, respectively) (57). Phage-bound D-IZN17 was rapidly captured from solution using Dynal T1 streptavidin-coated magnetic beads (Invitrogen) and briefly washed 3 times with 500 μl of 0.1% Tween 20 in Tris-buffered saline (wash buffer contained 100 μM D-biotin for the 1st wash). Phage was eluted in 50 μl of glycine (pH 2.2) elution buffer (10 min at RT) and neutralized with 7.5 μl of 1 M Tris, pH 9.1. The amplified phage library was sequenced prior to the first round of selection to confirm randomization, and preamplified eluted phage was sequenced following each round. All phage binding experiments were performed using the same protocol described above using 270 μM L-PIE2 soluble competitor. A 7-mer phage display was performed using a similar protocol.

Crystal growth and data collection. The original form of PIE12 (see Table 1) contains a C-terminal GK extension and did not yield highly diffracting crystals in complex with IQN17, a gp41 pocket mimic. Variants of PIE12 instead containing an N-terminal K or KG extension (K-PIE12, KHPCDYPEWQWLCEL; KG-PIE12, KGHPCDYPEWQWLCEL) crystallized in complex with IQN17 under a variety of conditions. In each case, the reservoir (850 μl) comprised a solution from a commercially available crystallization screen, and the crystallization drop was prepared by mixing 0.3 or 0.5 μl of the IQN17-PIE12 or IQN17-PIE71 protein solution (1:1.1 molar ratio, 10 mg/ml total in water) with 0.3 μl of the reservoir solution. Crystals typically grew in 1 to 10 days. All crystals were grown by sitting-drop vapor diffusion. IQN17-PIE12 form I crystals (KG-PIE12) were grown at 21°C in Hampton Scientific condition Screen II 48 (10% PEG 20,000, 0.1 M bicine, pH 9.0, 2% dioxane). IQN17-PIE12 form II crystals (KG-PIE12) were grown at 21°C in Emerald Biosystems condition Cryo-II 37 (50% ethylene glycol, 0.1 M imidazole, pH 8.0). IQN17-PIE12 form III crystals (K-PIE12) were grown at 4°C in Emerald Biosystems condition Cryo-II 25 (40% 2-methyl-2,4-pentanediol (MPD), 0.1 M *N*-cyclohexyl-3-aminopropanesulfonic acid (CAPS) [pH 10.5]). IQN17-PIE71 crystals were grown at 21°C in Qiagen PACT crystallization condition G4 (20% PEG 3350, 0.2 M potassium thiocyanate, 0.1 M bis-Tris propane, pH 7.5).

Crystals were mounted in a nylon loop and either directly cryocooled by plunging them into liquid nitrogen or cryocooled following brief (20 s) immersion in 20 μl crystallization buffer with 30% (IQN17-PIE12) or 15% (IQN17-PIE71) added glycerol. Crystals were maintained at 100 K during data collection. Data were collected either in the laboratory using a rotating copper anode X-ray generator or at a synchrotron beam line. Data were processed using the DENZO and SCALEPACK programs (40). All structures were determined by molecular replacement using the PHASER program (33) with IQN17-PIE7 as the search

model. The models were rebuilt using the O program (25) and refined against a maximum-likelihood target function using the REFMAC program (36). Structures were checked using the MolProbity program (6) (see Table 2 for data and refinement statistics).

Explanation of Lys placement. We were concerned that direct C-terminal addition of Lys would not be well tolerated because the D-peptide C-terminal region forms an α helix critically involved in the pocket-binding interface, with the C terminus itself being amidated for helix stability. Therefore, we inserted a Gly between the original C terminus of PIE7 and the C-terminal Lys, both to cap the helix and to separate the Lys from the binding interface. Unexpectedly, PIE7-GK-monomer is slightly more potent than PIE7 (see Table 1). A version of PIE7 containing an N- and C-terminal Lys (K-PIE7-GK) has the same potency as PIE7-GK (data not shown), indicating a beneficial effect imposed by the C-terminal Gly-Lys, as opposed to a deleterious effect created by a single Lys at the N terminus. This benefit is likely the reason that the linkage consisting of an ~22-Å cross-linker at the C terminus whose spacer arm consists of 5 PEGs (C_5C) results in a potency slightly superior to that of the N_5C linkage (see Table 1).

Viral infectivity assays. Pseudovirion infectivity assays were performed as described previously (57). Purified lyophilized inhibitors were dissolved in water (monomers) or 50 mM HEPES, pH 7.5 (dimers and trimers), to make high-concentration stocks. For HEPES-containing samples, all media were adjusted so that the HEPES content matched that in the sample with the highest HEPES concentration (typically, ~1 mM). HEPES at higher concentrations (e.g., 3 mM) enhanced infectivity up to ~15% but had minimal effect at ≤0.5 mM. The Monogram Biosciences PhenoSense Entry and peripheral blood mononuclear cell (PBMC) assays were performed as described previously (43, 52).

CD studies. Samples were prepared with 2 μM IZN17, a 1.1× molar ratio of inhibitor to target binding sites, phosphate-buffered saline (PBS; 50 mM sodium phosphate, 150 mM NaCl, pH 7.4), and 2 M GuHCl in a total volume of 2.5 ml. Thermal melts were performed by melting the sample in a square 1-cm cuvette from 25°C to 90°C (or 93°C for PIE12-trimer) in 2°C increments with 2 min of equilibration. To show reversibility, reverse melts were performed on each sample from 90°C to 30°C in 10°C increments with 5 min of equilibration. Data were averaged from a 30-s collection on an Aviv model 410 circular dichroism (CD) spectrapolarimeter.

For each sample, the CD data followed a smooth sigmoid transition as the sample was heated or cooled. The data were smoothed in the Kaleidagraph program (Synergy Software) using 2 points from both sides. The derivative value of the smoothed data was used to determine the point with the steepest rate of change on the melt curve, which is the melting temperature (T_m).

Passaging studies. Laboratory-adapted HIV-1 strain NL4-3 was generated by transient transfection of proviral DNA (pNL4-3) into 293T cells using Lipofectamine (Invitrogen). Cell-free supernatants containing virus were collected 48 h posttransfection and used to infect 5×10^5 CEM-1 cells in RPMI 1640 medium (0.5 ml). Virus was serially propagated once a week by 1:5 dilution of cell-free viral supernatants into fresh CEM-1 cells (5×10^5 cells, 0.5 ml) in the absence or presence of inhibitor (PIE7-dimer, PIE12-dimer, or PIE12-trimer). Viral titers were monitored biweekly by p24 antigen enzyme-linked immunosorbent assay (PerkinElmer). The inhibitor concentration started at approximately the 50% inhibitory concentrations (IC_{50}s; 20 nM for PIE7-dimer; 1 nM for PIE12-dimer, and PIE12-trimer) and was raised 1.5- to 2-fold when p24 antigen levels in inhibitor-containing cultures approached that in inhibitor-free cultures (usually 2 to 3 weeks for PIE7-dimer). PIE12-dimer and PIE12-trimer required a slower escalation strategy with prolonged incubation at a fixed inhibitor concentration for 5 to 15 weeks before escalation.

To identify PIE7-dimer escape mutations, viral RNA was isolated from cell-free supernatants of at least two cultures independently propagated in either the presence (resistant virus) or absence (control virus) of inhibitor (Qiagen RNA purification kit). Env cDNA was generated by reverse transcription (Eppendorf cMaster RTplus system and cMaster reverse transcription kit), amplified by PCR, and sequenced in five stretches (Thomas Jefferson University Nucleic Acid Facility). To confirm selected mutations in the gp41 N-peptide region, the cDNA segment encoding the gp41 ectodomain was reamplified by PCR and subcloned into the pAED4 vector, and the plasmid DNA from three or more individual clones was sequenced. The substitutions E560K and V570I were observed in all sequences from PIE7-dimer-resistant virus but were not observed in any sequence from control virus. An expression plasmid for HXB2 Env (pEBB_HXB2 Env) incorporating these substitutions was generated using site-directed mutagenesis (QuikChange; Stratagene) and was utilized in the pseudoviral infectivity assay described above.

Protein Data Bank accession numbers. The Protein Data Bank (PDB) accession numbers for the PIE12-IQN17 complex are 3L35, 3L36, and 3L37 for crystal forms I, II, and III, respectively, and 3MGN for the PIE71-IQN17 complex.

TABLE 1. D-peptide inhibition data[e]

Sample	Sequence	IC_{50} (nM)[a]	
		HXB2	JRFL
PIE7	KGA[PIE7]AA	620[b]	24,000[b]
PIE7-GK	GA[PIE7]AAGK	390	16,000
PIE7-GKK	GA[PIE7]AAGKK	380	19,000
PIE12	HP[PIE7]ELGK	37	580
PIE13[c]	HP[PIE7]KL	41	1,500
PIE14	HP[PIE7]RLGK	33	1,100
PIE15	HA[PIE7]ELGK	67	1,400
$N_9N(PIE7)_2$	$(KGA[PIE7]AA)_2$	1.9[b]	2,300[b]
$N_5C(PIE7)_2$	GA[PIE7]AAGKKGA[PIE7]AA	0.6	300
$C_5C(PIE7)_2$	$(GA[PIE7]AAGK)_2$	0.5	200
$C_5C(PIE12)_2$	$(HP[PIE7]ELGK)_2$	0.4	14
$N_9N(PIE7)_3$	$(KGA[PIE7]AA)_3$[d]	0.3[b]	220[b]
$C_5C(PIE7)_3$	$(GA[PIE7]AAGK)_3$[d]	0.1	6.7
$C_5C(PIE12)_3$	$(HP[PIE7]ELGK)_3$[d]	0.5	2.8
C37		1.4[b]	13[b]
Enfuvirtide		3.7[b]	5.0[b]

[a] The IC_{50} standard error of the mean is <25% for duplicate assays for all values.
[b] Values are from reference 57.
[c] PIE13 does not include a C-terminal GK extension because its C-terminal flanking sequence contains a Lys residue.
[d] The central peptide of each trimer has two tandem Lys residues (not shown).
[e] PIE7, CDYPEWQWLC, or PIE7 core motif.

RESULTS

Structure-guided phage display to optimize flanking residues. PIE inhibitors consist of a short core sequence surrounded by a disulfide bond that imparts structural rigidity required for binding (Table 1) (12). The large jump in affinity between our 1st-generation (12) and 2nd-generation (57) inhibitors was accomplished by optimizing this core sequence. There were also four fixed flanking residues outside the disulfide that arose from phage library cloning restrictions, Gly-Ala on the N terminus and Ala-Ala on the C terminus. Interestingly, our cocrystal structures of D-peptides in complex with a mimic of its gp41 pocket target (IQN17) reveal significant contacts between these presumed inert flanking residues and the pocket (12, 57). Thus, we reasoned that their optimization would likely lead to improved D-peptide affinity for the pocket.

To optimize these flanking residues, we used a commercially available phage library cloning system (NEB) that allowed us to relocate cloning sites away from the flanking regions (38). We redesigned the regions immediately outside the flanking residues in our cloning vector in order to structurally isolate them and minimize any bias caused by flanking sequence randomization. Using this vector, we constructed a phage library that varied only these four residues in the context of our previously optimized PIE7 core sequence (XXCDYPEWQW LCXX). After four rounds of panning, our phage library showed ~100-fold improved binding to a gp41 pocket mimic (D-IZN17) compared to that of clonal PIE7 phage with the original GA/AA flanking sequence. We extensively sequenced this phage pool to identify a consensus sequence, H(A/P)-[PIE7 core]-(R/K/E)L, as well as five dominant individual sequences. Using a phage clone binding assay, we found that these sequences bound the gp41 pocket 70- to 900-fold more tightly than PIE7, with PIE12 (HP-[PIE7 core]-EL) having the highest affinity (data not shown).

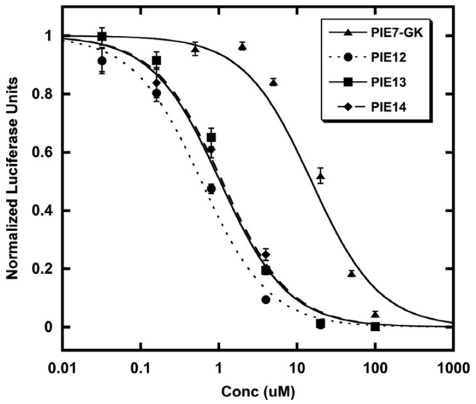


FIG. 1. Optimization of flanking residues enhances PIE potency. Each point represents the average of quadruplicate measurements from a representative pseudovirion entry inhibition assay (JRFL strain) normalized to the measurement for an uninhibited control. Error bars represent the standard errors of the means. PIE12 is ~2-fold more potent than PIE13 or PIE14 and is ~25-fold more potent than PIE7-GK.

Enhanced potency of 3rd-generation D-peptides. We synthesized D-peptides corresponding to the top three phage sequences in the binding assay (PIE12, PIE13, and PIE14) and tested their antiviral potencies in a pseudovirion entry assay (Table 1 and Fig. 1). Pairwise comparisons of both phage binding and inhibitor potency indicate that Pro is preferred over Ala at position 2 and Glu is preferred over Arg or Lys at position 13. As predicted from the phage binding assay, PIE12 has the best potency and is ~40-fold more potent than PIE7 (our best previously reported monomer) against strain JRFL.

Crystal structure of PIE12. To better understand the sources of PIE12's improved binding and potency, we crystallized PIE12 in complex with the N-trimer pocket mimic IQN17. Data were collected from three crystal forms (Table 2) at between 1.45- and 1.55-Å resolution. Each IQN17 trimer from the three crystal forms reported here and from the PIE7 structure (PDB accession number 2R5D) agreed well with one another (root mean square deviation [RMSD], 0.6 to 1.2 Å) on the basis of the least-squares overlap on all C_α atoms (residues 1 to 45 of all three chains). The structures suggest two sources of the improved affinity of PIE12 for IQN17 compared to that of PIE7. First, the new N-terminal flank residues (His1 and Pro2) form favorable ring stacking interactions with the pocket (IQN17-Trp571) (Fig. 2). Second, the substitution of Leu for Ala in the C-terminal flank sequence buries an additional ~50-Å^2 hydrophobic surface area in the pocket. Neither of these new interactions with the flanking sequence perturbs the original pocket-binding structure of the core PIE7 residues. Importantly, the structures reveal that PIE12's improved affinity does not result from new interactions with less conserved

TABLE 2. PIE12 and PIE71 crystallographic data and refinement statistics

Data	Result for PIE12 crystal:			Result for PIE71 crystal
	Form I	Form II	Form III	
Space group	$P2_1$	R3	P321	$P2_1$
Resolution (Å)	30.0–1.55 (1.61–1.55)[a]	30.0–1.45 (1.50–1.45)	30.0–1.45 (1.50–1.45)	30.0–1.40 (1.45–1.40)
No. of reflections measured	113,335	98,687	186,351	468,599
No. of unique reflections	25,088	10,475	14,802	82,774
Redundancy	4.5	9.4	12.6	5.7
Completeness (%)	86.5 (66.8)	97.1 (80.1)	99.7 (96.6)	98.2 (97.6)
$\langle I/\sigma I\rangle$[b]	18 (2.4)	19 (3.1)	17 (2.7)	15 (2.0)
Mosaicity (degree)	0.44	0.37	0.45	0.29
R_{sym}[c]	0.051 (0.250)	0.058 (0.102)	0.107 (0.235)	0.052 (0.316)
Refinement				
Resolution (Å)	30.0–1.55 (1.59–1.55)	30.0–1.45 (1.49–1.45)	30.0–1.45 (1.49–1.45)	30.0–1.40 (1.44–1.40)
No. of reflections used for refinement	23,765	9,448	13,629	80,532
No. of reflections in R_{free}[d] set	1,273	1,026	1,136	1,654
R_{cryst}[e]	0.232 (0.465)	0.234 (0.301)	0.243 (0.299)	0.261 (0.306)
R_{free}	0.288 (0.624)	0.264 (0.392)	0.278 (0.350)	0.288 (0.335)
RMSD bonds (Å)/angles (degrees)	0.012/1.440	0.013/1.693	0.010/1.530	0.009/1.094
<B>[g]				
All atoms (Å^2)/no. of atoms	23.7/1,172	31.9/384	29.2/384	Mol[f] 1, 24.3/1,555; mol 2, 36.0/1,491
PIE12 molecules only (Å^2)/no. of atoms	21.3/420	30.8/144	25.9/144	Mol 1, 18.3/368; mol 2, 39.9/322
Water molecules (Å^2)/no. of water atoms	32.0/197	38.0/36	40.6/49	39.9/389
ϕ/Ψ[h] most favored (%)	100	98.1	100	99.0

[a] Values in parentheses refer to data in the high-resolution shell.
[b] $\langle I/\sigma I\rangle$, average intensity of a group of reflections divided by the average standard deviation (sigma) of the same group of reflections.
[c] $R_{sym} = \Sigma|I - \langle I\rangle|/\Sigma I$, where I is the intensity of an individual measurement and $\langle I\rangle$ is the corresponding mean value.
[d] R_{free} is the same as R_{cryst} calculated with a randomly selected test set of reflections that were never used in refinement calculations.
[e] $R_{cryst} = \Sigma||Fo| - |Fc||/\Sigma|Fo|$, where $|Fo|$ is the observed and $|Fc|$ is the calculated structure factor amplitude.
[f] Mol, molecule.
[g] <B>, temperature factor.
[h] ϕ/ψ, dihedral angles.

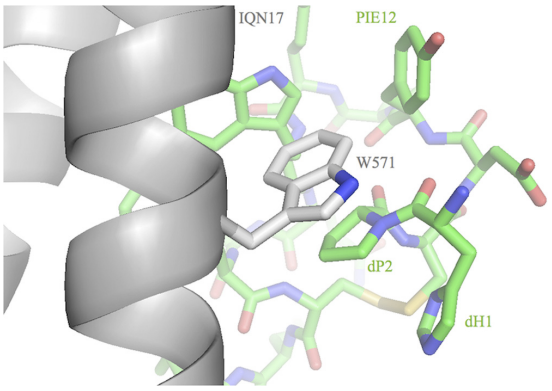


FIG. 2. Crystal structure of PIE12 binding to IQN17. Trp571 of the gp41 pocket (gray) and the N-terminal flank residues (dH1 and dP2) of PIE12 (green) appear to stabilize binding via ring-stacking interactions. The disulfide bond (yellow) is shown in the background.

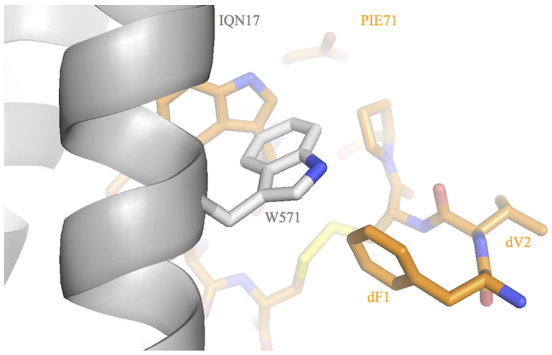


FIG. 3. Crystal structure of PIE71 binding to IQN17. The N-terminal flank residues (dF1 and dV2) of PIE71 (orange) are directed away from the pocket compared to the structure in PIE12 (Fig. 2). The disulfide bond (yellow) is shown in the background.

regions outside the pocket that might render PIE12 more vulnerable to resistance mutations.

Discovery and structure of a 7-mer D-peptide. The core sequence of PIE7 and PIE12 comprises 8 residues flanked by cysteines (8-mer). Modeling based on our 8-mer D-peptide/IQN17 crystal structures suggests that a 7-mer core is compatible with pocket binding of the WXWL consensus and formation of a disulfide bond (57). Previously, we saw that decreasing the size of the PIE core (from 10 to 8 residues) led to dramatically increased pocket binding (57), so we reasoned that further decreasing the size of the core might lead to additional potency gains. To explore this alternative geometry, we used a mirror-image discovery process similar to that employed with 8-mers to identify a 7-mer, PIE71 (FVCPPEWRWLCDL). PIE71 contains the same WXWL motif found in 8-mer and 10-mer pocket binders and inhibits strain HXB2 entry with an IC_{50} of 410 nM (data not shown), which is ~1.5 fold better than that of PIE7 but an order of magnitude worse than that of PIE12.

To gain a better understanding of the 7-mer binding solution, we determined a cocrystal structure of PIE71 in complex with IQN17 (Table 2). The key residues involved in the binding interface (WXWL) adopt nearly superposable conformations to those observed in PIE7 and PIE12, as do the C-terminal flank residues. However, the two structures deviate significantly at the N terminus (Fig. 2 and 3). Specifically, the 7-mer's disulfide bond is shifted much closer to the pocket, which directs the N-terminal flank residues away from the pocket region. As a result, the N-terminal flanking residues (Phe-Val) only graze the pocket, whereas PIE12's N-terminal flanking residues have an intimate interaction. So although the 7-mer is compatible with pocket binding, the smaller core is too constrained to allow optimal binding of the flank residues to the pocket. Due to this decreased binding interface and therefore decreased potency, we decided not to pursue the 7-mer geometry further.

Optimization of crosslinker length and geometry. We previously took advantage of the trimeric nature of the gp41 pocket target to geometrically increase the PIE7 binding affinity by cross-linking it into dimers and trimers (57). PIE7 has an N-terminal lysine, which furnishes a unique primary amino group (the N terminus is acetylated) and which was added for solubility. This lysine was used to produce dimers via reaction with a bis-PEG NHS ester crosslinker (NHS esters selectively react with primary amino groups). Trimers were produced by cross-linking two PIE7s to a central peptide with two lysines at the N terminus (2K-PIE7).

We hypothesized that the strength of the avidity effect is related to the length of the crosslinker and that shorter crosslinkers that still allow simultaneous binding to multiple pockets could strengthen potency. For the original N- to N-terminal linkage, we used a crosslinker with an ~35-Å spacer arm consisting of 9 PEG units (N_9N linkage). However, our crystal structures of D-peptides in complex with IQN17 reveal that C- to C-terminal or N- to C-terminal linkages could be significantly shorter and could be spanned by an ~22-Å crosslinker whose spacer arm consists of 5 PEGs (C_5C and N_5C linkages). Therefore, we relocated Lys to the C terminus of PIE7 (PIE7-GK) in order to make the N_5C heterodimers and C_5C homodimers (see Materials and Methods for additional details).

The resulting N_5C- and C_5C-PIE7-dimers have similar potencies that are significantly enhanced compared to the potency of our previous N_9N-PIE7-dimer (Table 1 and Fig. 4A). On the basis of these data, we chose C_5C connections as our standard linker, since they are simpler to produce than the hetero-N_5C linkage. Here, all dimers and trimers use the C_5C linkage unless otherwise specified. Combining our new optimized flanking residues and linkages, we produced PIE12-dimer and PIE12-trimer. Both are extremely potent against the difficult-to-inhibit primary strain JRFL (low-nanomolar IC_{50}s; Fig. 4B; Table 1), being up to 2 orders of magnitude more potent than our best previously described D-peptide (N_9N PIE7-trimer) (57).

Breadth against a diverse multiclade panel. HIV-1 has jumped from chimpanzees to humans at least three separate times, giving rise to groups M, N, and O (19). The main group (group M) accounts for >99% of all HIV-1 infections world-

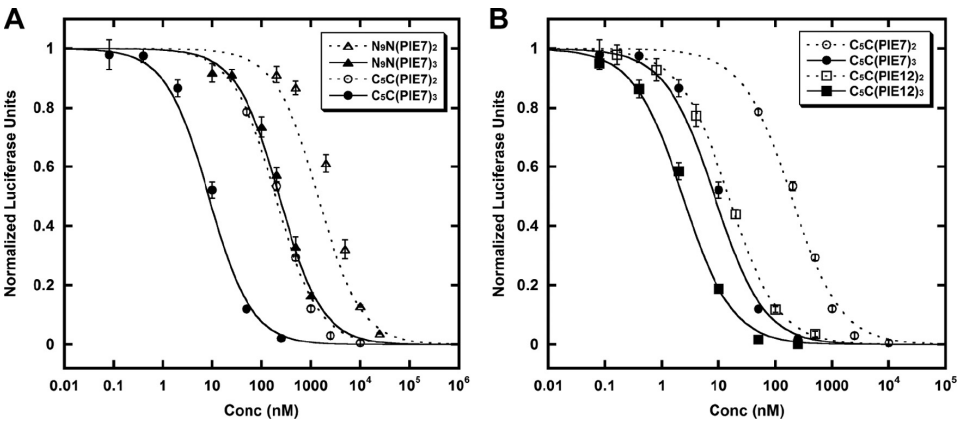


FIG. 4. Optimization of linkage geometry. Each point represents the average of quadruplicate measurements from a representative pseudovirion entry inhibition assay (JRFL strain) normalized to the measurement for the uninhibited control. Error bars represent the standard errors of the means. (A) Comparison of N_9N to C_5C linkages; (B) PIE7 versus PIE12-dimers and trimers (all C_5C linkages).

wide (32). HIV's high mutation rate has led to the emergence of diverse subtypes within group M that are categorized as clades A to D, F to H, J, and K and various circulating recombinant forms (CRFs; e.g., AE and BF). In 2000, clades A to D were estimated to represent >90% of HIV infections (39); however, in recent years CRFs have become more prevalent (1). Different subtypes contain up to 35% sequence diversity in Env, often causing antibodies raised against a particular strain to be ineffective against others (20).

To ensure that our pocket-specific D-peptides are potent and broadly neutralizing against the most common subtypes of HIV, we measured the potency of PIE7-trimer, PIE12-trimer, and PIE12 (with enfuvirtide as a control inhibitor) using the PhenoSense Entry pseudovirion assay (Monogram Biosciences) (Table 3) (43). The inhibitors were tested against a panel of 23 viruses pseudotyped with clonal and polyclonal envelopes representing clades A to D, several CRFs, and enfuvirtide-resistant strains. Both PIE7 and PIE12-trimers potently inhibited all strains tested, though PIE12-trimer was generally a superior inhibitor (and in all cases more potent than enfuvirtide). While PIE12-monomer is much less potent than PIE12-trimer, it is also broadly active. Interestingly, PIE12-trimer is ~10-fold more potent than PIE7-trimer against polyclonal virus from clades B and C (samples amplified from patient plasma), which is consistent with a resistance capacitor mechanism for maintaining potency in the presence of various Env sequences. All of the D-peptide inhibitors are unaffected by enfuvirtide resistance mutations. Additionally, lack of inhibition against a murine leukemia virus (MLV) control indicates that these inhibitors are specific and nontoxic in this assay.

Breadth against replication-competent primary viral isolates on PBMCs. To more closely mimic *in vivo* infection and further establish inhibitory breadth, we also tested the ability of PIE7-trimer, PIE12-trimer, and PIE12 to inhibit PBMC infection by replicating primary strains, again with enfuvirtide as a control (Table 4). These data confirm the potent and broad inhibitory activities of PIE7 and PIE12-trimer against all group M strains tested, including several CRFs. Toxicity was not observed on these cells at inhibitor concentrations up to 1 μM (the highest concentration tested), demonstrating a high therapeutic index for the trimers. Interestingly, the inhibitors are more potent in this assay than in the PhenoSense Entry assay, which may be due to differential receptor expression levels between the two cell types (45).

Notably, two group O strains were also tested in this assay and are much less sensitive to inhibition than group M strains. Group O contains several mutations (compared to the sequence of group M) in the pocket, including Q567R, T569S, K574R, Q577R, and V580L. The crystal structures of PIE7 and

TABLE 3. PhenoSense Entry assay data

HIV-1 isolate	Subtype	IC_{50} (nM)			
		PIE7-trimer	PIE12-trimer	PIE12-monomer	Enfuvirtide
A[a]	A	5.5	4.1	2,300	18
92RW008	A	2.0	1.0	1,400	7.2
92UG031	A	18	4.2	2,600	20
94KE105	AC	16	0.7	1,900	13
CMU02	AE	32	12	1,500	16
B[a]	B	140	13	3,300	30
1168	B	54	31	4,700	140.0
BaL	B	2.0	2.5	1,700	10
ENFr1[a]	B	2.0	0.8	790	760
ENFr2[a]	B	0.7	1.0	300	5,400
HXB2	B	0.1	0.3	50	2.6
JRCSF	B	13	3.4	1,100	14
JRFL	B	21	5.7	1,900	7.9
NL4.3	B	0.3	0.4	150	62
SF162	B	3.4	4.5	940	34
98CN009	BC	0.4	0.4	320	7.9
93BR029	BF	1.5	0.9	750	12
C[a]	C	220.0	26	5,100	71
97ZA012	C	2.0	0.7	1,500	10
98IN022	C	1.1	1.1	820	6.9
21068	C	6.6	5.0	1,800	47
D[a]	D	3.1	3.2	820	17
92UG005	D	3.9	2.5	2,000	10
aMLV		>10,000	>10,000	>500,000	>15,000

[a] Polyclonal viral envelopes amplified from patient plasma.

TABLE 4. PBMC assay data

HIV-1 isolate	Subtype	IC_{50} (nM)			
		PIE7-trimer	PIE12-trimer	PIE12-monomer	Enfuvirtide
92UG029	A	1.6	0.7	290	190
92UG037	A	0.1	0.2	36	41
93TH073	AE	0.6	0.8	270	200
CMU02	AE	0.2	0.4	300	44
CMU06	AE	0.3	0.4	210	5.7
IIIB	B	0.3	0.8	140	28
BaL	B	0.2	0.3	72	20
JRCSF	B	0.1	0.1	120	7.0
JRFL	B	0.5	0.3	110	1.7
93BR019	BF	1.7	4.7	170	>1,000
92BR025	C	15	5.2	>1,000	310
93IN101	C	0.4	0.4	160	22
92UG001	D	0.8	4.5	230	180
92UG046	D	0.1	1.2	170	130
93BR020	F	0.2	0.4	190	59
93BR029	F	0.2	0.8	86	19
G3	G	0.3	1.2	310	23
RU570	G	0.3	0.4	480	37
BCF01	Group O	>1,000	>1,000	>1,000	330
BCF02	Group O	>1,000	440	>1,000	0.4

PIE12 in complex with IQN17 reveal that, of these residues, the D-peptide directly interacts only with K574 (via a hydrophobic interaction) and Q577 (via hydrogen bonds). Group O gp41 has several other mutations in the groove just outside the pocket (i.e., H564E) that could also affect PIE potency (e.g., by slowing the association rate). It will be interesting to analyze the effects of these mutations in a group M (e.g., strain HXB2 or JRFL) background to see if they are responsible for the loss of potency.

Evidence for a charged resistance capacitor. With the design of PIE12-trimer, we now observe strong evidence for a highly charged resistance capacitor in which the PIE12-trimer pocket-binding affinity vastly exceeds the inhibitory potency. Comparing PIE7 and PIE12-trimers, we observe similar potencies against pseudovirion entry (Fig. 4B; Table 1), although we expect their target affinities to be extremely different.

Due to extraordinarily slow off rates, direct measurements of the pocket affinities for PIE7 and PIE12-trimers via surface plasmon resonance, used for earlier D-peptides (57), were not possible. Since the binding affinity of inhibitors correlates with the stability of inhibitor-target complexes, we used thermal denaturation monitored by CD to measure the relative stabilities of each IZN17-inhibitor complex and infer the relative affinities of our ultra-high-affinity binders. The melts were performed in 2 M GuHCl to destabilize the complexes and shift their melting points into an observable range (below 100°C).

The normalized thermal melts for each IZN17-inhibitor complex are plotted in Fig. 5, with T_m values being shown in the key. As expected, PIE12-trimer forms the most stable complex and has a T_m 8°C higher than that of the next most stable inhibitor complex (PIE7-trimer). PIE12 also forms a more stable complex than PIE7, as expected. Our previous experience showed that improvements in monomer affinity translated to approximately squared and cubed improvements in the corresponding dimers and trimers (57). On the basis of PIE12-trimer's optimized C_5C linkage (35-fold improved antiviral potency over that of the trimer with an N_9N linkage; strain JRFL data) and the ~25-fold difference in monomer potency between PIE7 and PIE12 (JRFL data), we estimate that PIE12-trimer binds to gp41 $>10^5$-fold (35×25^3) more tightly than N_9N PIE7-trimer. This predicted binding at subfemtomolar concentrations translates to a resistance capacitor charged to ~6 kcal/mol against strain JRFL. Interestingly, the potency plateau lies at a slightly better potency for trimers than for dimers, likely due to their faster association rates (i.e., three versus two opportunities for initial collision with the target).

Selection of resistant strains. To measure the resistance profile of our D-peptide inhibitors and test our resistance capacitor hypothesis, we conducted viral passaging studies with escalating inhibitor concentrations to select for resistant strains. These studies initially used PIE7-dimer, which was available from our previous study (57) and inhibits the parental strain, NL4-3, with an IC_{50} of ~20 nM. By doubling the PIE7-dimer concentration every 2 to 3 weeks, we obtained stable viral cultures in 2,000 nM inhibitor within 20 weeks of propagation. In comparison, we were able to obtain high-level enfuvirtide resistance (>1,000-fold) in only ~3 weeks using a similar protocol (H. K. Steger et al., submitted for publication).

Sequencing the N-peptide region of PIE7-dimer-resistant viruses revealed two selected mutations: E560K and V570I. These substitutions in the context of HXB2 pseudovirions conferred ~400-fold resistance to PIE7-dimer. These mutations also dramatically weaken the binding of D-peptides to the gp41 pocket but not the C-peptide inhibitor C37 (M. J. Root et al., unpublished data). It is not obvious from the PIE7 structure how these mutations weaken PIE7 binding. Despite this loss of affinity, the escape mutations had a minimal effect on the

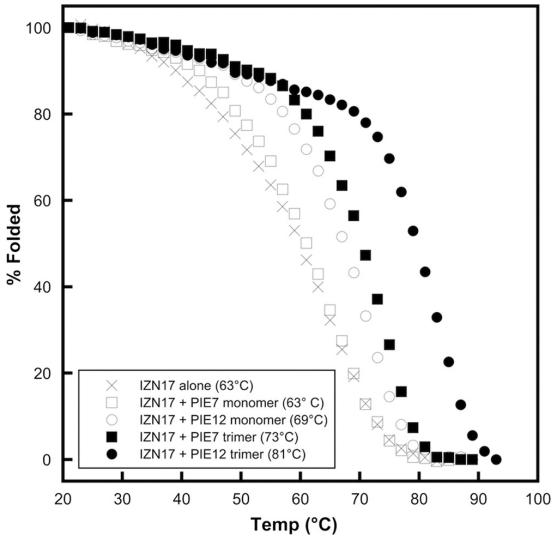


FIG. 5. Stability of D-peptide complexes. Normalized melting curves of IZN17 alone and with D-peptide inhibitors were monitored by CD in PBS–2 M GuHCl. T_m values are indicated in the key.

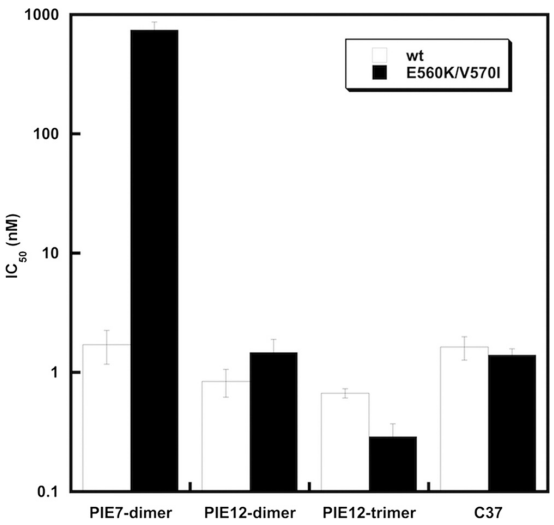


FIG. 6. Effect of PIE7-dimer resistance mutations on PIE7-dimer, PIE12-dimer, and PIE12-trimer potency. IC_{50}s against wild-type (wt) and PIE7-dimer-resistant (E560K/V570I) strain HXB2 pseudovirion entry are shown. The C-peptide inhibitor C37 is included as a control. Data represent the means from at least two independent experiments. Error bars represent the standard errors of the means.

potencies of PIE12-dimer and PIE12-trimer, as well as the C37 control inhibitor (Fig. 6). This result is predicted by the resistance capacitor hypothesis: affinity-disrupting escape mutations selected in the presence of weaker-binding inhibitors should be less disruptive to the potencies of tighter-binding inhibitors.

The rapid inhibitor escalation strategy utilized to generate PIE7-dimer resistance was not effective in generating HIV-1 resistant to PIE12-dimer or PIE12-trimer. Rather, the HIV-1 titer fell precipitously when inhibitor concentrations exceeded 20 nM (5 to 20 times the IC_{50}). Instead, we switched to a much slower escalation strategy with prolonged periods at stable inhibitor concentrations (5 to 15 weeks). Resistant virus emerged after 40 weeks of propagation in PIE12-dimer and after 65 weeks of propagation in PIE12-trimer. These observations suggest that a strong resistance capacitor profoundly delays selection of resistance mutations for these optimized fusion inhibitors.

Sequencing of the pocket region of PIE12-trimer-resistant viruses reveals only one mutation, Q577R. Interestingly, this substitution is present in nearly all group O isolates (including BCF01 and BCF02; Table 4) but is rare among group M isolates. Pseudovirions bearing Q577R confirm that this mutation confers substantial resistance to PIE12-trimer (data not shown). Examination of the PIE12 crystal structure shows that Q577 makes hydrogen bonds with Glu7 and Trp10 in PIE12, which may explain the disruptive effects of this mutation. Q577R's codon is predicted to disrupt the RRE stem-loop V structure, since it base pairs with the invariant W571 codon (Trp is encoded by only one codon).

DISCUSSION

PIE12-trimer is a D-peptide entry inhibitor with ~80-fold enhanced potency and an estimated >100,000-fold improved binding affinity compared to those of the best previously reported D-peptide. This dramatic improvement in affinity produces excellent breadth and a charged resistance capacitor to combat the emergence of resistance mutations. Indeed, PIE12-trimer was able to withstand the impact of resistance mutations to earlier D-peptides and required a much longer selection (65 weeks) to generate resistant strains. Ongoing work is exploring the mechanism of PIE7-dimer, PIE12-dimer, and PIE12-trimer resistance and its relationship to group O's insensitivity. A key question is whether HIV can develop resistance to these inhibitors independent of changes in affinity (e.g., kinetics) that are capable of maintaining viral fitness.

Viral escape affects even the newest class of FDA-approved HIV-1 drugs, integrase inhibitors. Resistance to raltegravir and corresponding treatment failure were observed in a significant subset of patients in both the phase II and III clinical studies (5), and corresponding resistance mutations can be seen within 4 weeks when resistant virus is selected in viral passaging studies (28). Our studies indicate that PIE12-trimer is a promising entry inhibitor that could overcome the limitations associated with the two currently approved entry inhibitors, enfuvirtide (high dosing, susceptibility to resistance) and maraviroc (Selzentry; effective only against R5 viruses) and may also prove to have a better resistance profile than even the newest class of HIV-1 inhibitors.

In addition to being a possible therapeutic agent, PIE12-trimer is an ideal candidate for a topical microbicide, as its protease resistance would allow it to withstand the protease-rich environment of the vaginal mucosa. In the absence of a safe and effective HIV vaccine, a topical microbicide to prevent the sexual transmission of HIV is an urgent unmet global health need. The ultimate utility of PIE12-trimer as a microbicide or therapeutic agent will be determined by advanced preclinical and clinical studies, including characterization of pharmacokinetics, *in vivo* toxicity, effectiveness in animal models of HIV infection (alone or in combination with other HIV inhibitors), and optimization of formulations for microbicide gels or vaginal rings.

More generally, the present work unequivocally shows that D-peptide inhibitors can be designed with high potency and specificity against natural L-protein targets. The D-peptide design methodology described here can be applied to diverse biomedical applications, particularly for the many viruses that share HIV's hairpin-closing entry mechanism (e.g., influenza virus, Ebola virus, respiratory syncytial virus, severe acute respiratory syndrome coronavirus, Dengue virus, and West Nile virus). Our resistance capacitor design strategy may also be generally applicable for treating other rapidly evolving diseases, especially when combined with recent advances in anticipating likely structural sources of drug resistance (37). Finally, the development of PIE12-trimer as a strong clinical candidate will allow D-peptide therapeutics to be evaluated *in vivo* to determine if their theoretical advantages warrant a prominent role as a new class of therapeutic agents.

ACKNOWLEDGMENTS

We thank Bob Schackmann and Scott Endicott (University of Utah Peptide Synthesis Core Facility) for peptide synthesis, Yu Shi for early 7-mer phage display, and Dong Han and Pham Phung (Monogram) for technical assistance with the PhenoSense Entry assay.

PBMC assays were performed by Southern Research Institute (principal investigator, Roger Ptak), funded by contract HHSN272200700041C (from the National Institute of Allergy and Infectious Diseases, National Institutes of Health [NIH], U.S. Department of Health and Human Services). This work was supported by grants from the NIH to M.S.K. (AI076168), M.J.R. (GM066682), and C.P.H. (GM082545), as well as a University of Utah Technology Commercialization Grant to M.S.K. Portions of this research were carried out at the Stanford Synchrotron Radiation Lightsource (SSRL), a national user facility operated by Stanford University on behalf of the Office of Basic Energy Sciences, U.S. Department of Energy. The SSRL Structural Molecular Biology Program is supported by the Office of Biological and Environmental Research, U.S. Department of Energy, and by the National Center for Research Resources, Biomedical Technology Program, NIH, and the National Institute of General Medical Sciences.

B.D.W., D.M.E., and M.S.K. are cofounders of Kayak Biosciences. This startup company is focused on advancing D-peptide inhibitors to the clinic.

REFERENCES

1. **Buonaguro, L., M. L. Tornesello, and F. M. Buonaguro.** 2007. Human immunodeficiency virus type 1 subtype distribution in the worldwide epidemic: pathogenetic and therapeutic implications. J. Virol. **81:**10209–10219.
2. **Chan, D. C., C. T. Chutkowski, and P. S. Kim.** 1998. 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. **95:**15613–15617.
3. **Chan, D. C., D. Fass, J. M. Berger, and P. S. Kim.** 1997. Core structure of gp41 from the HIV envelope glycoprotein. Cell **89:**263–273.
4. **Chinnadurai, R., D. Rajan, J. Munch, and F. Kirchhoff.** 2007. Human immunodeficiency virus type 1 variants resistant to first- and second-version fusion inhibitors and cytopathic in ex vivo human lymphoid tissue. J. Virol. **81:**6563–6572.
5. **Cooper, D. A., R. T. Steigbigel, J. M. Gatell, J. K. Rockstroh, C. Katlama, P. Yeni, A. Lazzarin, B. Clotet, P. N. Kumar, J. E. Eron, M. Schechter, M. Markowitz, M. R. Loutfy, J. L. Lennox, J. Zhao, J. Chen, D. M. Ryan, R. R. Rhodes, J. A. Killar, L. R. Gilde, K. M. Strohmaier, A. R. Meibohm, M. D. Miller, D. J. Hazuda, M. L. Nessly, M. J. DiNubile, R. D. Isaacs, H. Teppler, and B. Y. Nguyen.** 2008. Subgroup and resistance analyses of raltegravir for resistant HIV-1 infection. N. Engl. J. Med. **359:**355–365.
6. **Davis, I. W., A. Leaver-Fay, V. B. Chen, J. N. Block, G. J. Kapral, X. Wang, L. W. Murray, W. B. Arendall III, J. Snoeyink, J. S. Richardson, and D. C. Richardson.** 2007. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res. **35:**W375–W383.
7. **Debnath, A. K., L. Radigan, and S. Jiang.** 1999. Structure-based identification of small molecule antiviral compounds targeted to the gp41 core structure of the human immunodeficiency virus type 1. J. Med. Chem. **42:**3203–3209.
8. **Derdeyn, C. A., J. M. Decker, J. N. Sfakianos, Z. Zhang, W. A. O'Brien, L. Ratner, G. M. Shaw, and E. Hunter.** 2001. Sensitivity of human immunodeficiency virus type 1 to fusion inhibitors targeted to the gp41 first heptad repeat involves distinct regions of gp41 and is consistently modulated by gp120 interactions with the coreceptor. J. Virol. **75:**8605–8614.
9. **Dwyer, J. J., K. L. Wilson, D. K. Davison, S. A. Freel, J. E. Seedorff, S. A. Wring, N. A. Tvermoes, T. J. Matthews, M. L. Greenberg, and M. K. Delmedico.** 2007. Design of helical, oligomeric HIV-1 fusion inhibitor peptides with potent activity against enfuvirtide-resistant virus. Proc. Natl. Acad. Sci. U. S. A. **104:**12772–12777.
10. **Eckert, D. M., and P. S. Kim.** 2001. Design of potent inhibitors of HIV-1 entry from the gp41 N-peptide region. Proc. Natl. Acad. Sci. U. S. A. **98:**11187–11192.
11. **Eckert, D. M., and P. S. Kim.** 2001. Mechanisms of viral membrane fusion and its inhibition. Annu. Rev. Biochem. **70:**777–810.
12. **Eckert, D. M., V. N. Malashkevich, L. H. Hong, P. A. Carr, and P. S. Kim.** 1999. Inhibiting HIV-1 entry: discovery of D-peptide inhibitors that target the gp41 coiled-coil pocket. Cell **99:**103–115.
13. **Eggink, D., C. E. Baldwin, Y. Deng, J. P. Langedijk, M. Lu, R. W. Sanders, and B. Berkhout.** 2008. Selection of T1249-resistant human immunodeficiency virus type 1 variants. J. Virol. **82:**6678–6688.
14. **Ernst, J. T., O. Kutzki, A. K. Debnath, S. Jiang, H. Lu, and A. D. Hamilton.** 2002. Design of a protein surface antagonist based on alpha-helix mimicry: inhibition of gp41 assembly and viral fusion. Angew. Chem. Int. ed. Engl. **41:**278–281.
15. **Ferrer, M., T. M. Kapoor, T. Strassmaier, W. Weissenhorn, J. J. Skehel, D. Oprian, S. L. Schreiber, D. C. Wiley, and S. C. Harrison.** 1999. Selection of gp41-mediated HIV-1 cell entry inhibitors from biased combinatorial libraries of non-natural binding elements. Nat. Struct. Biol. **6:**953–960.
16. **Freed, E. O., and M. A. Martin.** 1995. The role of human immunodeficiency virus type 1 envelope glycoproteins in virus infection. J. Biol. Chem. **270:**23883–23886.
17. **Frey, G., S. Rits-Volloch, X. Q. Zhang, R. T. Schooley, B. Chen, and S. C. Harrison.** 2006. Small molecules that bind the inner core of gp41 and inhibit HIV envelope-mediated fusion. Proc. Natl. Acad. Sci. U. S. A. **103:**13938–13943.
18. **Furuta, R. A., C. T. Wild, Y. Weng, and C. D. Weiss.** 1998. Capture of an early fusion-active conformation of HIV-1 gp41. Nat. Struct. Biol. **5:**276–279.
19. **Gao, F., E. Bailes, D. L. Robertson, Y. Chen, C. M. Rodenburg, S. F. Michael, L. B. Cummins, L. O. Arthur, M. Peeters, G. M. Shaw, P. M. Sharp, and B. H. Hahn.** 1999. Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes. Nature **397:**436–441.
20. **Gaschen, B., J. Taylor, K. Yusim, B. Foley, F. Gao, D. Lang, V. Novitsky, B. Haynes, B. H. Hahn, T. Bhattacharya, and B. Korber.** 2002. Diversity considerations in HIV-1 vaccine selection. Science **296:**2354–2360.
21. **Golding, H., M. Zaitseva, E. de Rosny, L. R. King, J. Manischewitz, I. Sidorov, M. K. Gorny, S. Zolla-Pazner, D. S. Dimitrov, and C. D. Weiss.** 2002. Dissection of human immunodeficiency virus type 1 entry with neutralizing antibodies to gp41 fusion intermediates. J. Virol. **76:**6780–6790.
22. **Jiang, S., K. Lin, N. Strick, and A. R. Neurath.** 1993. HIV-1 inhibition by a peptide. Nature **365:**113.
23. **Jiang, S., H. Lu, S. Liu, Q. Zhao, Y. He, and A. K. Debnath.** 2004. N-substituted pyrrole derivatives as novel human immunodeficiency virus type 1 entry inhibitors that interfere with the gp41 six-helix bundle formation and block virus fusion. Antimicrob. Agents Chemother. **48:**4349–4359.
24. **Jin, B. S., J. R. Ryu, K. Ahn, and Y. G. Yu.** 2000. Design of a peptide inhibitor that blocks the cell fusion mediated by glycoprotein 41 of human immunodeficiency virus type 1. AIDS Res. Hum. Retroviruses **16:**1797–1804.
25. **Jones, T. A., J. Y. Zou, S. W. Cowan, and M. Kjeldgaard.** 1991. Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr. A **47**(Pt 2)**:**110–119.
26. **Judice, J. K., J. Y. Tom, W. Huang, T. Wrin, J. Vennari, C. J. Petropoulos, and R. S. McDowell.** 1997. Inhibition of HIV type 1 infectivity by constrained alpha-helical peptides: implications for the viral fusion mechanism. Proc. Natl. Acad. Sci. U. S. A. **94:**13426–13430.
27. **Kahle, K. M., H. K. Steger, and M. J. Root.** 2009. Asymmetric deactivation of HIV-1 gp41 following fusion inhibitor binding. PLoS Pathog. **5:**e1000674.
28. **Kobayashi, M., K. Nakahara, T. Seki, S. Miki, S. Kawauchi, A. Suyama, C. Wakasa-Morimoto, M. Kodama, T. Endoh, E. Oosugi, Y. Matsushita, H. Murai, T. Fujishita, T. Yoshinaga, E. Garvey, S. Foster, M. Underwood, B. Johns, A. Sato, and T. Fujiwara.** 2008. Selection of diverse and clinically relevant integrase inhibitor-resistant human immunodeficiency virus type 1 mutants. Antiviral Res. **80:**213–222.
29. **Lalezari, J. P., N. C. Bellos, K. Sathasivam, G. J. Richmond, C. J. Cohen, R. A. Myers, Jr., D. H. Henry, C. Raskino, T. Melby, H. Murchison, Y. Zhang, R. Spence, M. L. Greenberg, R. A. Demasi, and G. D. Miralles.** 2005. T-1249 retains potent antiretroviral activity in patients who had experienced virological failure while on an enfuvirtide-containing treatment regimen. J. Infect. Dis. **191:**1155–1163.
30. **Lalezari, J. P., K. Henry, M. O'Hearn, J. S. Montaner, P. J. Piliero, B. Trottier, S. Walmsley, C. Cohen, D. R. Kuritzkes, J. J. Eron, Jr., J. Chung, R. DeMasi, L. Donatacci, C. Drobnes, J. Delehanty, and M. Salgo.** 2003. Enfuvirtide, an HIV-1 fusion inhibitor, for drug-resistant HIV infection in North and South America. N. Engl. J. Med. **348:**2175–2185.
31. **Lazzarin, A., B. Clotet, D. Cooper, J. Reynes, K. Arasteh, M. Nelson, C. Katlama, H. J. Stellbrink, J. F. Delfraissy, J. Lange, L. Huson, R. DeMasi, C. Wat, J. Delehanty, C. Drobnes, and M. Salgo.** 2003. Efficacy of enfuvirtide in patients infected with drug-resistant HIV-1 in Europe and Australia. N. Engl. J. Med. **348:**2186–2195.
32. **Marx, P. A.** 2005. Unsolved questions over the origin of HIV and AIDS. ASM News **71:**15–20.
33. **McCoy, A. J., R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, and R. J. Read.** 2007. Phaser crystallographic software. J. Appl. Crystallogr. **40:**658–674.
34. **Miller, M. D., R. Geleziunas, E. Bianchi, S. Lennard, R. Hrin, H. Zhang, M. Lu, Z. An, P. Ingallinella, M. Finotto, M. Mattu, A. C. Finnefrock, D. Bramhill, J. Cook, D. M. Eckert, R. Hampton, M. Patel, S. Jarantow, J. Joyce, G. Ciliberto, R. Cortese, P. Lu, W. Strohl, W. Schleif, M. McElhaugh, S. Lane, C. Lloyd, D. Lowe, J. Osbourn, T. Vaughan, E. Emini, G. Barbato, P. S. Kim, D. J. Hazuda, J. W. Shiver, and A. Pessi.** 2005. A human monoclonal antibody neutralizes diverse HIV-1 isolates by binding a critical gp41 epitope. Proc. Natl. Acad. Sci. U. S. A. **102:**14759–14764.
35. **Munoz-Barroso, I., S. Durell, K. Sakaguchi, E. Appella, and R. Blumenthal.** 1998. Dilation of the human immunodeficiency virus-1 envelope glycoprotein fusion pore revealed by the inhibitory action of a synthetic peptide from gp41. J. Cell Biol. **140:**315–323.
36. **Murshudov, G. N., A. A. Vagin, and E. J. Dodson.** 1997. Refinement of

macromolecular structures by the maximum-likelihood method. Acta Crystallogr. D Biol. Crystallogr. **53:**240–255.

37. **Nalam, M. N., A. Ali, M. D. Altman, G. S. Reddy, S. Chellappan, V. Kairys, A. Ozen, H. Cao, M. K. Gilson, B. Tidor, T. M. Rana, and C. A. Schiffer.** 2010. Evaluating the substrate-envelope hypothesis: structural analysis of novel HIV-1 protease inhibitors designed to be robust against drug resistance. J. Virol. **84:**5368–5378.
38. **Noren, K. A., and C. J. Noren.** 2001. Construction of high-complexity combinatorial phage display peptide libraries. Methods **23:**169–178.
39. **Osmanov, S., C. Pattou, N. Walker, B. Schwardlander, and J. Esparza.** 2002. Estimated global distribution and regional spread of HIV-1 genetic subtypes in the year 2000. J. Acquir. Immune Defic. Syndr. **29:**184–190.
40. **Otwinowski, Z., and W. Minor.** 1997. Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. **276:**307–326.
41. **Pappenheimer, J. R., C. E. Dahl, M. L. Karnovsky, and J. E. Maggio.** 1994. Intestinal absorption and excretion of octapeptides composed of D amino acids. Proc. Natl. Acad. Sci. U. S. A. **91:**1942–1945.
42. **Pappenheimer, J. R., M. L. Karnovsky, and J. E. Maggio.** 1997. Absorption and excretion of undegradable peptides: role of lipid solubility and net charge. J. Pharmacol. Exp. Ther. **280:**292–300.
43. **Petropoulos, C. J., N. T. Parkin, K. L. Limoli, Y. S. Lie, T. Wrin, W. Huang, H. Tian, D. Smith, G. A. Winslow, D. J. Capon, and J. M. Whitcomb.** 2000. A novel phenotypic drug susceptibility assay for human immunodeficiency virus type 1. Antimicrob. Agents Chemother. **44:**920–928.
44. **Ray, N., J. E. Harrison, L. A. Blackburn, J. N. Martin, S. G. Deeks, and R. W. Doms.** 2007. Clinical resistance to enfuvirtide does not affect susceptibility of human immunodeficiency virus type 1 to other classes of entry inhibitors. J. Virol. **81:**3240–3250.
45. **Reeves, J. D., S. A. Gallo, N. Ahmad, J. L. Miamidian, P. E. Harvey, M. Sharron, S. Pohlmann, J. N. Sfakianos, C. A. Derdeyn, R. Blumenthal, E. Hunter, and R. W. Doms.** 2002. Sensitivity of HIV-1 to entry inhibitors correlates with envelope/coreceptor affinity, receptor density, and fusion kinetics. Proc. Natl. Acad. Sci. U. S. A. **99:**16249–16254.
46. **Reeves, J. D., F. H. Lee, J. L. Miamidian, C. B. Jabara, M. M. Juntilla, and R. W. Doms.** 2005. Enfuvirtide resistance mutations: impact on human immunodeficiency virus envelope function, entry inhibitor sensitivity, and virus neutralization. J. Virol. **79:**4991–4999.
47. **Rimsky, L. T., D. C. Shugars, and T. J. Matthews.** 1998. Determinants of human immunodeficiency virus type 1 resistance to gp41-derived inhibitory peptides. J. Virol. **72:**986–993.
48. **Root, M. J., M. S. Kay, and P. S. Kim.** 2001. Protein design of an HIV-1 entry inhibitor. Science **291:**884–888.
49. **Sadowski, M., J. Pankiewicz, H. Scholtzova, J. A. Ripellino, Y. Li, S. D. Schmidt, P. M. Mathews, J. D. Fryer, D. M. Holtzman, E. M. Sigurdsson, and T. Wisniewski.** 2004. A synthetic peptide blocking the apolipoprotein E/beta-amyloid binding mitigates beta-amyloid toxicity and fibril formation in vitro and reduces beta-amyloid plaques in transgenic mice. Am. J. Pathol. **165:**937–948.
50. **Schumacher, T. N., L. M. Mayr, D. L. Minor, Jr., M. A. Milhollen, M. W. Burgess, and P. S. Kim.** 1996. Identification of D-peptide ligands through mirror-image phage display. Science **271:**1854–1857.
51. **Sia, S. K., P. A. Carr, A. G. Cochran, V. N. Malashkevich, and P. S. Kim.** 2002. Short constrained peptides that inhibit HIV-1 entry. Proc. Natl. Acad. Sci. U. S. A. **99:**14664–14669.
52. **Southern Research Institute.** 2008, posting date. Anti-HIV evaluation assays in fresh human cells. Southern Research Institute, Birmingham, AL. http://www.southernresearch.org/contract-services/anti-hiv-evaluation-assays.html#fresh_human_cells.
53. **Steger, H. K., and M. J. Root.** 2006. Kinetic dependence to HIV-1 entry inhibition. J. Biol. Chem. **281:**25813–25821.
54. **Stephens, O. M., S. Kim, B. D. Welch, M. E. Hodsdon, M. S. Kay, and A. Schepartz.** 2005. Inhibiting HIV fusion with a beta-peptide foldamer. J. Am. Chem. Soc. **127:**13126–13127.
55. **Tan, K., J. Liu, J. Wang, S. Shen, and M. Lu.** 1997. Atomic structure of a thermostable subdomain of HIV-1 gp41. Proc. Natl. Acad. Sci. U. S. A. **94:**12303–12308.
56. **Weissenhorn, W., A. Dessen, S. C. Harrison, J. J. Skehel, and D. C. Wiley.** 1997. Atomic structure of the ectodomain from HIV-1 gp41. Nature **387:**426–430.
57. **Welch, B. D., A. P. VanDemark, A. Heroux, C. P. Hill, and M. S. Kay.** 2007. Potent D-peptide inhibitors of HIV-1 entry. Proc. Natl. Acad. Sci. U. S. A. **104:**16828–16833.
58. **Wild, C., J. W. Dubay, T. Greenwell, T. Baird, Jr., T. G. Oas, C. McDanal, E. Hunter, and T. Matthews.** 1994. Propensity for a leucine zipper-like domain of human immunodeficiency virus type 1 gp41 to form oligomers correlates with a role in virus-induced fusion rather than assembly of the glycoprotein complex. Proc. Natl. Acad. Sci. U. S. A. **91:**12676–12680.
59. **Wild, C., T. Greenwell, and T. Matthews.** 1993. A synthetic peptide from HIV-1 gp41 is a potent inhibitor of virus-mediated cell-cell fusion. AIDS Res. Hum. Retroviruses **9:**1051–1053.
60. **Wild, C., T. Oas, C. McDanal, D. Bolognesi, and T. Matthews.** 1992. A synthetic peptide inhibitor of human immunodeficiency virus replication: correlation between solution structure and viral inhibition. Proc. Natl. Acad. Sci. U. S. A. **89:**10537–10541.
61. **Wild, C. T., D. C. Shugars, T. K. Greenwell, C. B. McDanal, and T. J. Matthews.** 1994. Peptides corresponding to a predictive alpha-helical domain of human immunodeficiency virus type 1 gp41 are potent inhibitors of virus infection. Proc. Natl. Acad. Sci. U. S. A. **91:**9770–9774.

CHAPTER 4

SHORT-CIRCUITING THE RESISTANCE CAPACITOR: HIV RESISTANCE MECHANISM AGAINST D-PEPTIDE ENTRY INHIBITORS

Matthew T. Weinstock, Amanda E. Siglin, Amanda R. Smith, Ethan B. Howell, Frank G. Whitby, J. Nicholas Francis, Christopher P. Hill, Michael J. Root, Michael S. Kay

Introduction

In 2012, an estimated 35.3 million individuals were living with HIV, a virus that has resulted in an estimated 36 million AIDS-related deaths since the beginning of the epidemic (UNAIDS report on the global AIDS epidemic 2013 fact sheet).

The advent of therapeutics targeting various stages of the viral lifecycle has dramatically improved the clinical outcome for those with access to treatment; however, resistance has been observed to develop against all classes of approved drugs (1). This resistance is due to a highly error-prone viral reverse transcriptase, frequent recombination events, and rapid turnover of the virus (2, 3).

We recently reported the development of cost-effective and extremely potent D-peptide HIV-1 entry inhibitors through mirror-image phage display (4-6) and the further improvement of potency to low pM levels by developing a crosslinking scaffold allowing for both trimerization of the peptide and attachment of a membrane targeting moiety that pre-positions the inhibitor at sites of viral entry (7).

These peptides act by tightly binding to a hydrophobic pocket at the base of the HIV-1 gp41 N-peptide trimeric coiled coil, and act by preventing the collapse of N- and C-peptide regions of gp41 to a 6-helix bundle structure, a structural transition required to bring the cellular and viral membranes in apposition for fusion (8) (see Fig. 2.1).

Because resistance against other entry inhibitors with similar modes of action (e.g., the FDA-approved Fuzeon) has been observed (9, 10), we were interested to evaluate the resistance profile of our lead candidates PIE12-trimer and cholesterolated PIE12-trimer (chol-PIE12-trimer).

We previously reported a high barrier to resistance relative to other entry inhibitors (e.g., Fuzeon and previous generation D-peptides) for PIE12-trimer (5), which we hypothesize is the result of two distinctive features of our most potent D-peptides. First, while both Fuzeon and PIE12-trimer inhibit HIV entry by binding to the gp41 N-peptide region, Fuzeon binds to an extended groove while PIE12-trimer binds to a separate deep hydrophobic pocket at the base of the groove (11). This hydrophobic pocket is interesting in that it plays a key role in mediating viral entry (12) and the pocket-encoding region of the Env gene also contains the Rev response element (RRE), a structured RNA element required for nuclear export of viral RNA. The hydrophobic pocket displays a remarkable degree of sequence conservation likely reflecting that mutations in this region must simultaneously preserve functionality of the fusion machinery and the RRE, thus restricting the range of potential resistance mutations and thereby raising the barrier to resistance against pocket-targeting entry inhibitors.

Second, during the optimization of previous generation D-peptide inhibitors, we noticed that as binding affinity for the pocket was improved via successive rounds of phage display, we reached a plateau in potency (6). We proposed that this potency plateau is due to the transient nature of the gp41 pre-hairpin intermediate conformation targeted by these compounds. Because the pre-hairpin intermediate conformation exists for a short time window, there is a point where association kinetics rather than binding affinity becomes the limiting factor for therapeutic potency. Despite this plateau in potency, we continued to "over-engineer" the D-peptide binding affinity for the pocket. We reasoned that this excess of binding energy would sever the link between drug potency and affinity for the pocket such that a mutation that subtly weakens binding

affinity will not affect potency, and thus will not be provided with the selective advantage necessary to spread through the viral population. We hypothesized that this "resistance capacitor" would prevent the development of drug resistance through pathways involving the incremental accumulation of subtle affinity-disrupting mutations.

While a high barrier to resistance was observed, we nonetheless reported the selection of strains resistant to PIE12-dimer and PIE12-trimer through viral passaging with escalating inhibitor concentrations (5). Herein we extend the same methodology to cholesterolated PIE12-trimer (chol-PIE12-trimer). Furthermore, we comprehensively characterize the genotype of PIE12-dimer, PIE12-trimer, and chol-PIE12-trimer resistant viral populations through various sequencing methodologies, describe the biochemical and structural elucidation of a primary mechanism of resistance against PIE12-trimer and cholesterolated PIE12-trimer, discuss our progress towards characterizing the fitness and level of resistance of the resistant viral pools, and outline our strategy to overcome resistance.

Materials and Methods

Passaging studies

PIE12-dimer and PIE12-trimer resistant viruses were acquired in a previous study **(5)**. chol-PIE12-trimer resistant viruses were generated starting from two different viral stocks. Laboratory-adapted HIV-1 NL4-3 was passaged in the presence of escalating chol-PIE12-trimer concentrations as follows: 0.1 nM for 12 weeks, 0.4 nM for 4 weeks, and 1.6 nM for 5 weeks. A second population of chol-PIE12-trimer resistant viruses was generated from PIE12-trimer resistant NL4-3 (that had been propagating in 320 nM PIE12-trimer) by passaging in the presence of escalating chol-PIE12-trimer

concentrations as follows: 30 nM for 12 weeks, 120 nM for 3 weeks, 480 nM for 4 weeks, and 1000 nM for 5 weeks.

Generation of envelope cDNA

Viral RNA was isolated with the E.Z.N.A.® Viral RNA Kit (Omega Bio-Tek). cDNA for the envelope gene was generated using the SuperScript® III One-Step RT-PCR System with Platinum® *Taq* High Fidelity (Invitrogen). Primers were designed using Primer3Plus (13).

Sanger sequencing of Env clones

Sanger sequencing of viral envelope was conducted either as previously described (5) or as follows. Envelope cDNA was cloned into various standard vectors using standard restriction enzyme cloning or TOPO® TA cloning (Life Technologies). Vectors were transformed into *E. coli* and plated out on LB agar plates with appropriate antibiotic resistance. Plasmids were isolated (GenElute™ Plasmid Miniprep Kit, Sigma) from cultures inoculated with individual colonies and were submitted to the DNA Sequencing Core Facility (University of Utah) for Sanger sequencing of the viral envelope.

Deep sequencing and bioinformatic analysis

Envelope cDNA from resistant populations was purified by SDS-PAGE, extracted (GenElute™ Gel Extraction Kit, Sigma), and submitted to the Microarray and Genomic Analysis Core Facility (University of Utah) for library preparation and sequencing. Barcoded samples were multiplexed onto a single lane and sequenced on an Illumina HiSeq 2000 (50 cycle single-end reads). The raw sequencing reads in FASTQ format were mapped to nucleotides 6195-8826 from the HIV-1 vector pNL4-3 sequence

(GenBank: AF324493.2) using STAMPY (14). A summary of base calls for each position of the reference sequence was generated by converting the alignment data to pileup format using the SAMtools:*mpileup* function (used -BQ0 and -d 2000000 options). Custom perl scrips (available upon request) were used to convert base calls at each position from pileup format to a more intuitive table, and then to filter the data to display only those positions where deviations from the reference are present in the population above a user-defined threshold (10%).

Peptide synthesis

All synthetic peptides were acetylated at the N-terminus and amidated at the C-terminus. All reverse-phase HPLC (RP-HPLC) purifications were performed on a Waters C18 column (BEH X-Bridge, 19 x 250 mm, 10 μm, 300Å) with water/acetonitrile gradient in 0.1% TFA unless otherwise noted. KG-PIE12 was synthesized by the University of Utah DNA/Peptide Core as described in (6) and purified via RP-HPLC on a Vydac C18 column (TP, 22 x 250 mm, 10 μm, 300Å). PIE12-GK, PIE12-trimer, and cholesterol-PIE12-trimer were synthesized as previously described (5). PIE12-Tide Fluor™ 2 was prepared by reacting PIE12-GK (~1.5 mM) with 1.5x excess of Tide Fluor™ 2 succinimidyl ester (AAT Bioquest) in dry DMAC with 250 mM triethanolamine. The reaction was run overnight at room temp in the dark. The reaction was quenched by the addition of AcOH to 5% and then purified via RP-HPLC. SelMet-IQN17-Q577R and HIV C34 were synthesized on a PS3 peptide synthesizer (Protein Technologies, Inc.) at 80 umol scale on NovaPEG Rink Amide LL resin (Novabiochem) and Rink Amide AM resin LL (Novabiochem), respectively, via Fmoc-SPPS and were purified by RP-HPLC.

Recombinant protein production

Cys-Gly-Gly-Asp-IZN36 (derived from HXB2 strain) cloned into pET14b behind the His tag and thrombin cleavage site had been prepared previously (6). A Q577R mutant version was generated via site-directed mutagenesis using the QuikChange® protocol (Stratagene). Protein expression was performed in BL21 (DE3) cells (Novagen) following Studier's autoinduction protocol (15). Cultures were grown in Fernbach flasks shaking at 150 RPM overnight at 37ºC. Cells were pelleted by centrifugation, resuspended in 20 mM sodium phosphate pH 8.0, 300 mM NaCl, 10 mM Imidazole, and frozen at -80ºC until further use. Cells were disrupted by sonication and centrifuged for 20 minutes at 20,000xg (4ºC). The pellet (containing IZN36 in inclusion bodies) was solubilized in 20 mM sodium phosphate pH 8.0, 300 mM NaCl, 10 mM Imidazole, 6 M GuHCl, and sonicated again. The sample was clarified by centrifugation (20 min, 20,000xg) and applied to His-Select Ni-affinity resin (Sigma). Following thorough washing, the protein was eluted in 20 mM sodium phosphate pH 8.0, 300 mM NaCl, 250 mM Imidazole, 6 M GuHCl. Samples were dialyzed into H2O with 0.1% TFA and purified via RP-HPLC on a Waters C18 column (BEH X-Bridge, 19 x 250 mm,10 μm, 300Å) with water/acetonitrile gradient in 0.1% TFA and lyophilized. The cysteine residues of these constructs were biotinylated for SPR. Reactions were carried out in PBS pH 7.0 with ~1-2mg/mL protein and a 15x molar excess of Biotin-dPEG®3-MAL (Quanta BioDesign) at room temp for 1 hr. Proteins were purified via RP-HPLC as above. Removal of the His tag was carried out in PBS buffer pH 6.3 with 10 uM heparin at a protein concentration of ~0.8mg/mL with 1 unit Thrombin protease (GE Healthcare)

per mg of protein. The cleavage reaction was incubated at 37ºC with agitation for 44 hrs followed by RP-HPLC as above.

Protein interaction analysis

Protein interaction analysis was carried out on a Biacore 3000. Streptavidin (Pierce) was immobilized on CM5 chips (GE Healthcare) using the Amine Coupling Kit (GE Healthcare) following a standard coupling protocol (16). Biotinylated IZN36 constructs (WT and Q577R mutant) were captured on the surface to a density of between 100-200 RUs. IZN36 stocks for target loading (~50 nM) were prepared in running buffer in the presence of 1.2x molar excess of HIV C34 (we found that loading IZN36 in the presence of C34 results in better behaved surfaces). C34 was removed by surface regeneration with 0.1% SDS injections. A free, Streptavidin-derivatized surface served as a blank. Analyses were performed at 25ºC in sterile filtered and degassed Gibco® PBS buffer pH 7.4 (Invitrogen) supplemented with 0.005% P20 and 1mg/mL BSA at a flow rate of 50 μL/min. Stock solutions of PIE12GK were prepared at 50 μM in running buffer from which samples in the concentration series were generated by serial dilution. Injections were performed for 1 min using KINJECT. Dissociation was monitored for 2 min, followed by a 20 μL injection (QUICKINJECT) of running buffer followed by EXTRACLEAN. Sensorgrams were processed using Scrubber2 (BioLogic Software), and data were analyzed with Prism 5.0 (GraphPad Software). The equilibrium binding data were fit using the built-in saturation binding model (One site -- Specific binding), normalized to Bmax, and replotted as concentration of analyte vs. % capacity.

Crystallography

KG-PIE12 was crystallized in complex with IQN17-Q577R (1:1.1 IQN17:KG-PIE12, 10 mg/mL total in water) under a variety of conditions. Crystal trays were prepared in Intelli-Plate 96-3 well plates (Art Robbins Instruments) using a Crystal Phoenix (Art Robbins Instruments). In each case, the reservoir (60 μL) consisted of a commercially available crystallization screen and the crystallization drop (0.6 μL) contained either 2:1, 1:1, or 1:2 protein:reservoir condition. Crystals typically grew in less than 10 days by sitting-drop vapor diffusion. IQN17-Q577R crystals were grown at 21°C in Hampton Research condition SaltRx HT B4 (1.8 M ammonium citrate dibasic, 0.1 M sodium acetate trihydrate, pH 4.6). Crystals were mounted in a nylon loop and either directly cryocooled by plunging into liquid nitrogen or cryocooled following a 20 s immersion in 20 μL crystallization buffer with 20% added glycerol. Crystals were maintained at 100 K during data collection. Data collection was performed at the Stanford Synchrotron Radiation Lightsource (SSRL). Data were processed using DENZO and SCALEPACK (17). The IQN17-Q577R/PIE12 structure was determined by molecular replacement using PHASER (18) with PIE12-IQN17 (pdb structure 3L37) as the search model. The model was rebuilt using O (19) and COOT (20) and refined against a maximum-likelihood target function using PHENIX (21). The structure was checked using the MolProbity program (22). See Fig. 4.1 for crystallographic data.

Infectivity of the polyclonal resistant pools

Infectivity of polyclonal resistant pools was measured using TZM-BL cells following standard protocols (our collaborator (MJR) is currently preparing a detailed protocol for the publication of this work).

Results

Passaging

We previously reported the development of HIV-1 populations resistant to PIE12-dimer and PIE12-trimer derived from a parental laboratory-adapted NL4-3 strain by passaging the virus in the presence of escalating concentrations of inhibitor; acquiring PIE12-dimer resistant virus propagating in 1280 nM PIE12-dimer (~85x IC_{50} of parental strain) and PIE12-trimer resistant virus propagating in 320 nM PIE12-trimer (~750x IC_{50} of parental strain) (5). Similar passaging studies were performed with the inhibitor chol-PIE12-trimer beginning with either naive virus (laboratory-adapted HIV-1 NL4-3) or PIE12-trimer resistant virus. Starting with naive NL4-3, we were able to acquire a virus able to stably propagate in the presence of 1.6 nM chol-PIE12-trimer (~80x the IC_{50} of the parental strain) after 21 weeks of passaging. By starting the experiment with a resistant strain of NL4-3 able to passage in the presence of 320 nM PIE12-trimer (well 1), we were able to acquire a virus able to propagate in the presence of 1000 nM chol-PIE12-trimer (~50,000x the IC_{50} of the parental strain) after 24 weeks of passaging.

Inhibition of resistant pools

In order to characterize the degree of resistance acquired during the passaging, inhibition assays were performed on the polyclonal resistant viral pools to calculate the new IC_{50} values. PIE12-dimer resistant virus, a series of archived viruses from the PIE12-trimer resistance passaging (taken from different time points when the virus was passaging at different concentrations of inhibitor - 10, 40, 80, 320 nM), and both chol-PIE12-trimer resistant populations (one generated starting from naive NL4-3 with the other starting from PIE12-trimer resistant virus (propagating at 320 nM PIE12-trimer,

well 1)), were all assayed. While the initial data presented in Fig. 4.2 are preliminary (and will be repeated with more precision), it is clear that high levels of resistance were acquired for all inhibitors.

Several interesting observations can be made from the inhibition data. 1) While PIE12-dimer resistance does impact PIE12-trimer inhibition, PIE12-trimer is still able to inhibit the PIE12-dimer resistant population at low nM concentrations; however, PIE12-trimer resistance completely abolishes the potency of PIE12-dimer. This suggests two separate resistance mechanisms, with the PIE12-trimer resistance mechanism being more severe. 2) Cross-resistance to C37 (an entry inhibitor that works through a similar mechanism) was only observed in one population (PIE12-dimer resistant). This suggests that while a more general resistance mechanism is at play (e.g., altered fusion kinetics) in the PIE12-dimer resistant virus, a more specific mechanism is being utilized by PIE12-trimer and chol-PIE12-trimer resistant viruses that does not necessarily affect other entry inhibitors targeting the N-peptide region. 3) While chol-PIE12-trimer resistant virus (started from the naive NL4-3 parental strain) displays similar resistance against both chol-PIE12-trimer and PIE12-trimer compared to the PIE12-trimer resistant viruses, the chol-PIE12-trimer virus (started from the PIE12-trimer resistant virus (well 1)) manifests a marked increase in resistance to both PIE12-trimer and chol-PIE12-trimer.

Sequencing

To thoroughly characterize the genetic basis of resistance, viral RNA was isolated from PIE12-trimer, PIE12-dimer, and chol-PIE12-trimer resistant viral pools as well as RNA from a control virus that had been propagated in the absence of any inhibitors. The isolated RNA was used to generate corresponding cDNA of the Env gene, which was

either cloned into a vector for sequencing of full-length Env clones, or submitted for deep sequencing. These two sequencing strategies were pursued in parallel to leverage the relative strengths of each method.

The advantage of sequencing the entire Env gene from single clones via Sanger sequencing is that linkage information from mutations that reside anywhere in the genome is preserved; however, the drawback is that due to the time, labor, and cost involved in acquiring sequences of Env clones in this manner, it is relatively low throughput. While deep sequencing is not able to yield long-range linkage information between potential linked mutations (as the reads are very short), it does give a remarkable depth of coverage of Env genes from a resistant population (we achieved an average depth of coverage of ~260,000x for each position of Env for each sample submitted), giving a comprehensive view of genetic space.

We submitted Env cDNA from laboratory-adapted HIV-1 NL4-3 (to serve as a control to help distinguish between laboratory adaptations and resistance mutations in our resistant populations); the PIE12-dimer resistant virus able to propagate in the presence of 1280 nM PIE12-dimer; the series of archived viruses from the PIE12-trimer passaging study from different time points while it was passaging at various levels of PIE12-trimer (10, 40, 80, and 320 nM PIE12-trimer); another PIE12-trimer resistant virus propagating at 320 nM inhibitor from an independent passaging experiment (denoted as well 2 to distinguish it from the 320 nM sample from the aforementioned series (which is denoted well 1)); the chol-PIE12-trimer resistant virus (started from naive NL4-3) able to propagate at 1.6 nM chol-PIE12-trimer; and the chol-PIE12-trimer resistant virus (started

from PIE12-trimer resistant virus propagating at 320 nM PIE12-trimer) able to propagate at 1000 nM chol-PIE12-trimer.

Sanger sequencing data on a total of 13 PIE12-trimer resistant clones (five clones from virus propagating at 320 nM in well 1, and eight clones from the virus propagating at 320 nM in well 2) was also performed, and the data are in excellent agreement with the deep sequencing analysis (data not shown).

218 positions in the sequenced region across all resistant populations (PIE12-dimer, PIE12-trimer, chol-PIE12-trimer) were identified in the deep sequencing data where the sum of all nonreference calls (including insertions and deletions) was greater than 10% of the total calls. For preliminary analysis, we reasoned that 10% should be a high enough threshold that most of the noise due to genetic drift will be filtered out, but low enough to catch minor variations in the population that could give rise to resistance.

Mutations have been further filtered to include only those mutations that occur at high prevalence in a population (≥50%), are not silent (unless in RRE where a silent mutation could affect RNA structure), occur at any frequency at an interesting location in Env, or occur at any frequency when present at the same position across multiple independent samples, resulting in a table of 71 total positions mutated across all sequenced samples (28, PIE12-dimer; 13, PIE12-trimer (10 nM); 14, PIE12-trimer (40 nM); 15, PIE12-trimer (80 nM); 21, PIE12-trimer (320 nM, well 2); 20, PIE12-trimer (320 nM, well 1); 26, chol-PIE12-trimer (started from PIE12-trimer resistant virus); 30, chol-PIE12-trimer (started from naive NL4-3)).

A few interesting features stand out that might hint at important resistance mutations (or important compensatory mutations associated with resistance mutations).

There are two prominent in-frame deletions (Δ509-520/Δ512-523 in the V1/V2 loop of gp120, and Δ1206-1220 in the V4 loop of gp120) that occur in all of the resistant populations except PIE12-dimer (lacks the V1/V2 loop deletion) and the 10 nM PIE12-trimer (lacks the V4 loop deletion). The V1/V2 loop deletion occurs in all populations with high PIE12-trimer resistance, indicating that it is likely causing or is associated with (i.e., compensatory mutation) high PIE12-trimer resistance. The V4 loop mutation occurs at a prevalence of ~50% in populations that contain it; however, the region deleted is immediately adjacent to an identical stretch of sequence. Thus, it is possible that the deletion is present at ~100% in the populations that contain it, and that the reported 50% is a bioinformatic artifact. Future analysis will determine if this is the case.

There are many mutations in residues at glycosylation sites (canonical NXT/S sites) that will modify the glycosylation pattern of Env. These modifications could potentially affect its structure or activity (e.g., fusion kinetics) in such a way that fosters resistance.

There are a few mutations in the CD4-binding interface that could affect how Env interacts with its receptor.

The mutation of residue Q577 (to R, N, or K) in all of the PIE12-trimer and chol-PIE12-trimer resistant populations is of special interest. Previous preliminary sequence analysis conducted on a limited number of clones from a PIE12-trimer resistant population focusing on the N-peptide region of Env identified the presence of the Q577R mutation in the pocket region, a residue that directly interacts with the PIE12 peptide (5). Subsequent work has shown that Q577R impacts the potency of both PIE12-trimer (IC_{50} changes from 0.72 to > 3 μM) and chol-PIE12-trimer (IC_{50} changes from 0.013 to 10.1

nM) when incorporated into HXB2 pseudovirions (7). A similar effect has been observed for the Q577K mutation (data not shown). This pseudovirion data, taken with the above polyclonal inhibition data, suggest that Q577R is a key determinant of resistant in PIE12-trimer and chol-PIE12-trimer populations.

Interestingly, the Q577R mutation was not seen in PIE12-dimer resistant populations, where there is a different mutation in the binding site (I580L). Because this mutation is conservative, is not directly involved in the PIE12/pocket interaction, and has been shown to have no effect on potency of PIE12-trimer or chol-PIE12-trimer (data not shown), its role in PIE12-dimer resistance appears unlikely, but awaits further examination. It appears that the main resistance mutations in the PIE12-dimer resistant population lie far from the binding pocket, a phenomenon not unheard of as mutations outside the binding site have been associated with reduced susceptibility to entry inhibitors (2).

Because the chol-PIE12-trimer resistant population (started from naive NL4-3) displays a similar level of resistance to PIE12-trimer and chol-PIE12-trimer as both the PIE12-trimer resistant populations and the HXB2 Q577R pseudovirions (mentioned above), and because this chol-PIE12-trimer resistant population contains ~100% Q577R, it is probable that Q577R alone accounts for the majority of the chol-PIE12-trimer resistance phenotype in this population.

Because resistance against chol-PIE12-trimer increased (IC_{50} changed from ~10 nM to > 100 nM) in the passaging studies started with PIE12-trimer resistant virus (320 nM, well 1), and because there are relatively few sequence differences between these two populations, it is possible to narrow the mutations responsible for this increase in

resistance against chol-PIE12-trimer by comparing the sequences and noting mutations unique to the chol-PIE12-trimer resistant population. 13 differences occur in the filtered data set (F175L, V200E, S306R, M326L, G354E, S387T, N460I, H643Y, N674T, P727L, S810R, V829I, and V832L), two of which are at glycosylation sites (S387T and N674T), one is in the C-peptide region (H643Y), and one (M326L) is in the V3 loop and has been implicated in being involved in the gp120/co-receptor interface (23).

Protein interaction analysis of Q577R

The crystal structure of PIE12 in complex with an N-peptide pocket mimic indicates that Q577 makes important hydrogen bond contacts with residues in PIE12 (Glu7 and Trp10) (5), and their disruption via the Q577R mutation is likely to weaken the binding affinity of the D-peptide, resulting in the resistance seen in the PIE12-trimer and chol-PIE12-trimer resistant populations and the previously mentioned HXB2 pseudovirions bearing only the Q577R mutation (see discussion above). To determine the impact of the Q577R on PIE12 binding affinity, the mutation was incorporated into an N-peptide mimic (IZN36 (24)) and the interaction was analyzed via surface plasmon resonance analysis (SPR). The mutation results in a dramatic weakening of the interaction (Fig. 4.3). Furthermore, because our inhibitor is trimeric, we expect the diminished binding affinity seen for the monomer to have a more pronounced effect on the observed binding affinity of PIE12-trimer. Similar results were obtained through fluorescence polarization using a fluorescently labeled PIE12 peptide (data not shown). These data, coupled with the sequencing and inhibition data for the polyclonal resistant pools (and the HXB2 Q577R pseudovirion data), indicates that Q577R (and likely Q577N and Q577K seen in the deep sequencing data) is a major cause of resistance

against PIE12-trimer and chol-PIE12-trimer. This also likely explains the high level of PIE12-trimer resistance previously seen in Group O strains, as the isolates we have assayed contain Q577R (as well as a number of other pocket mutations (5)).

Crystallographic analysis of Q577R

To further interrogate the effect of Q577R on the PIE12/pocket interaction, monomeric PIE12 was crystallized in complex with an HIV-1 N-peptide pocket mimic (IQN17 (24)) bearing the Q577R mutation. An overlay of the region around position 577 of the acquired structure of PIE12/IQN17 (Q577R) with a structure of PIE12/IQN17 (wt) (pdb structure 3L37) is presented along with crystallographic statistics in Fig. 4.1. As expected, the Q577R mutation disrupts hydrogen bonds between Q577 and two PIE12 residues (Glu7 and Trp10). There is also a repositioning of two adjacent residues (residue Glu9 in PIE12 and L561 in IQN17).

Discussion and Future Directions

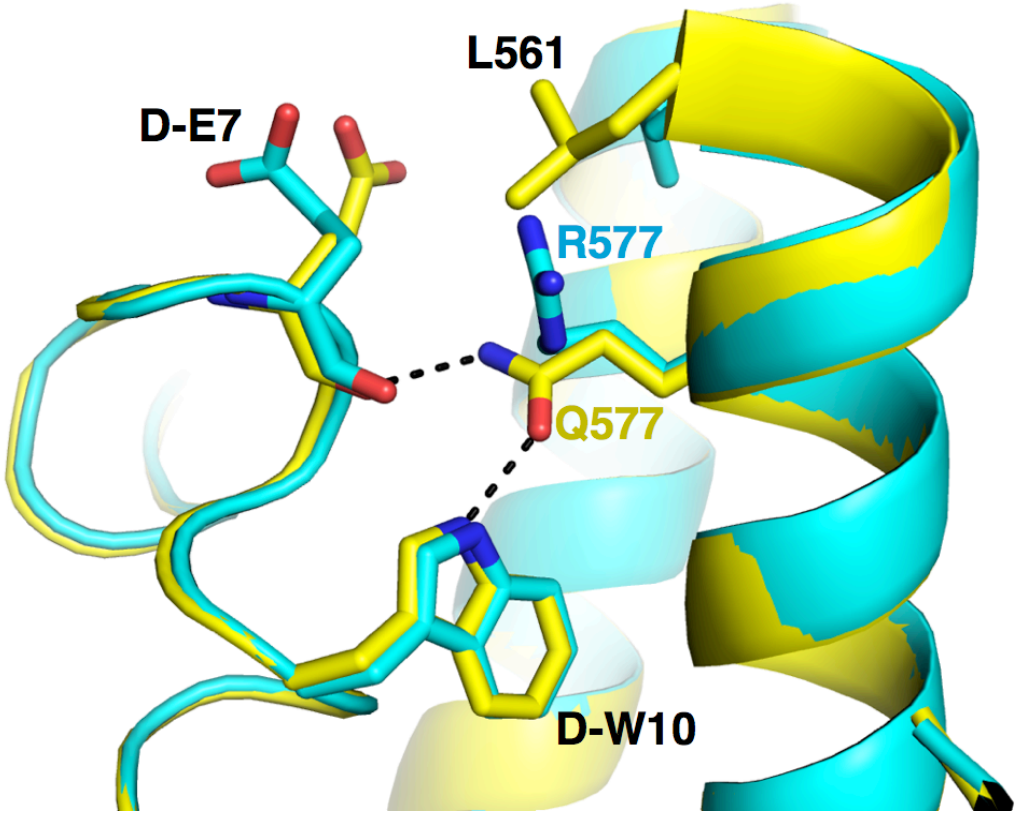
In summary, by applying genomic, biophysical, virological, and structural methods, we have determined that Q577R is a major cause of resistance to PIE12-containing D-peptide entry inhibitors. One remaining question is why Q577R is so difficult to acquire (i.e., why do these peptides have such a high barrier to resistance if one mutation can dramatically diminish their potency?). One possibility is that while Q577R is sufficient to impart significant resistance, it might negatively impact fitness. Thus, in order for Q577R to establish itself, it might require compensatory mutations to repair the fitness defect. In line with this hypothesis, while preliminary fitness data show no obvious signs of a serious growth defect in any of the resistant populations (data not

shown), a prior mutagenesis study on the gp41 N-peptide demonstrated that while pseudovirions can be generated with the Q577R mutation, and that Env in packaged and processed to wt levels, the virions exhibit a severe infectivity defect (infectivity < 5% of wt) (25). Studies are already underway to characterize the fitness of the polyclonal resistant pools, as well as the Q577R/N/K mutations alone in a wt background. If a fitness defect is observed, we will determine the compensatory mutations that restore fitness by either adding more mutations into the wt background, or starting with an Env clone from a resistant population containing the compensatory mutations and subtracting them out until a fitness defect is observed.

We are also interested in determining the mutations that result in PIE12-dimer and high chol-PIE12-trimer resistance, and if anything other than Q577R/N/K contributes to PIE12-trimer resistance. To this end we are currently generating Env expression plasmids containing Env clones from the resistant populations. For Env sequences from the PIE12-trimer resistant populations, we have reset Q577R back to wt (to see if any other mutations contribute to resistance). By generating chimeric Env constructs (e.g., wt gp120, mutant gp41, or vice versa) and then narrowing down from there, we hope to determine mutations responsible for resistance.

Ultimately we are interested in not simply understanding resistance, but overcoming it. To this end, we have already cloned Q577R into a target for mirror-image phage display, and efforts are underway to discover novel D-peptide entry inhibitors that bind with high affinity to the mutant pocket, or that bind well to both the mutant and wt pocket.

A)



B)

Crystallographic Data and Refinement Statistics

SelMet-IQN17 Q577R / KG-PIE12	
Space Group	R_3
Contents per asymmetric unit	1 IQN monomer, 1 D-peptide
Resolution	1.3 Å
R_{crys} [1]	0.197
R_{free} [2]	0.225

[1] $R_{crys}=\sum||Fo|-|Fc||/\sum|Fo|$

[2] R_{free} is the same as R_{crys} calculated with a randomly selected test set of reflection that were never used in refinement.

Figure 4.1. Crystallographic analysis of Q577R mutation. A) Overlay of crystal structures of PIE12 binding to the wt pocket (yellow) and a mutant (Q577R) pocket (turquoise) showing the disruption of important hydrogen bonding interactions mediated by residue 577. B) Data and Refinement statistics.

	PIE12-dimer (nM)	PIE12-trimer (nM)	chol-PIE12-trimer (nM)	C37 (nM)
control NL4-3	15	0.43	0.02	3.4
PIE12-dimer resistant (1280 nM)	860	52	-	30
PIE12-trimer resistant (40 nM)	>1,000	~2000	~10	8.5
PIE12-trimer resistant (80 nM)	-	~2000	~10	4
PIE12-trimer resistant (320 nM)	-	~2000	~10	5
chol-PIE12-trimer resistant (from naive)	-	~2000	~10	7.5
chol-PIE12-trimer resistant (from PIE12-trimer resistant virus)	-	>> 1000	>100	6.9

Figure 4.2. Inhibitory data for polyclonal resistant pools. Preliminary IC_{50} values for polyclonal resistant pools. Yellow indicates a modest level of resistance (< 10x change in IC_{50}), light red indicates high level resistance (> 50x change in IC_{50}), and dark red indicates an extreme level of resistance (> 2,000x change in IC_{50})

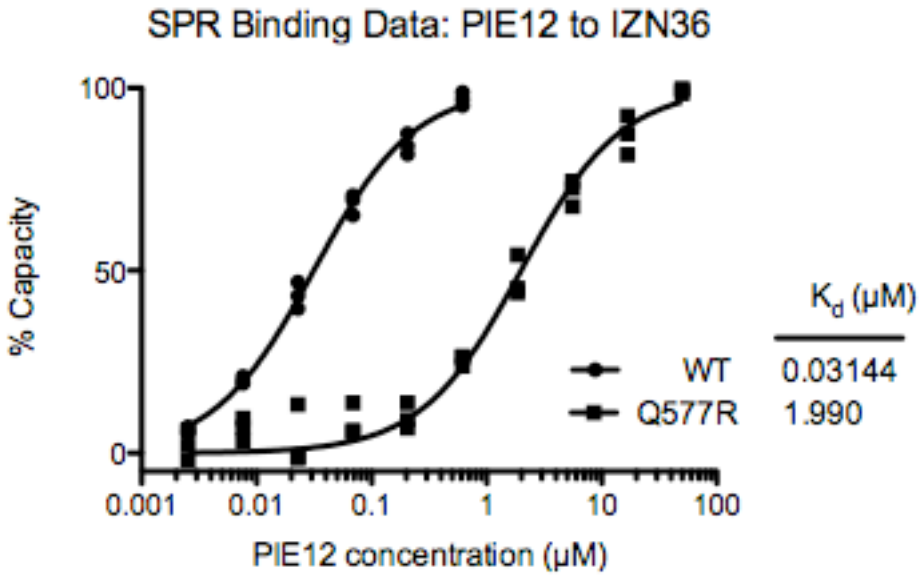


Figure 4.3. Protein interaction analysis of Q577R mutation. Equilibrium binding data (in triplicate) for interaction between immobilized IZN36 (wt or Q577R) and PIE12 monomer.

References

1. Menendez-Arias L. Molecular basis of human immunodeficiency virus drug resistance: an update. Antiviral research. 2010;85(1):210-31. doi: 10.1016/j.antiviral.2009.07.006. PubMed PMID: 19616029.

2. Clavel F, Hance AJ. HIV drug resistance. The New England journal of medicine. 2004;350(10):1023-35. doi: 10.1056/NEJMra025195. PubMed PMID: 14999114.

3. Rambaut A, Posada D, Crandall KA, Holmes EC. The causes and consequences of HIV evolution. Nat Rev Genet. 2004;5(1):52-61. doi: Doi 10.1038/Nrg1246. PubMed PMID: ISI:000187641100016.

4. Weinstock MT, Francis JN, Redman JS, Kay MS. Protease-resistant peptide design-empowering nature's fragile warriors against HIV. Biopolymers. 2012;98(5):431-42. doi: 10.1002/bip.22073. PubMed PMID: 23203688; PubMed Central PMCID: PMC3548907.

5. Welch BD, Francis JN, Redman JS, Paul S, Weinstock MT, Reeves JD, et al. Design of a potent D-peptide HIV-1 entry inhibitor with a strong barrier to resistance. J Virol. 2010;84(21):11235-44. doi: 10.1128/JVI.01339-10. PubMed PMID: 20719956; PubMed Central PMCID: PMC2953169.

6. Welch BD, VanDemark AP, Heroux A, Hill CP, Kay MS. Potent D-peptide inhibitors of HIV-1 entry. Proc Natl Acad Sci U S A. 2007;104(43):16828-33. Epub 2007/10/19. doi: 0708109104 [pii] 10.1073/pnas.0708109104. PubMed PMID: 17942675; PubMed Central PMCID: PMC2040420.

7. Francis JN, Redman JS, Eckert DM, Kay MS. Design of a Modular Tetrameric Scaffold for the Synthesis of Membrane-Localized d-Peptide Inhibitors of HIV-1 Entry. Bioconjug Chem. 2012. doi: 10.1021/bc300076f. PubMed PMID: 22545664; PubMed Central PMCID: PMC3429785.

8. Eckert DM, Kim PS. Mechanisms of viral membrane fusion and its inhibition. Annu Rev Biochem. 2001;70:777-810. Epub 2001/06/08. doi: 70/1/777 [pii] 10.1146/annurev.biochem.70.1.777. PubMed PMID: 11395423.

9. Eggink D, Langedijk JP, Bonvin AM, Deng Y, Lu M, Berkhout B, et al. Detailed mechanistic insights into HIV-1 sensitivity to three generations of fusion inhibitors. J Biol Chem. 2009;284(39):26941-50. doi: 10.1074/jbc.M109.004416. PubMed PMID: 19617355; PubMed Central PMCID: PMC2785381.

10. De Feo CJ, Weiss CD. Escape from human immunodeficiency virus type 1 (HIV-1) entry inhibitors. Viruses. 2012;4(12):3859-911. PubMed PMID: 23342377; PubMed Central PMCID: PMC3528295.

11. Chan DC, Fass D, Berger JM, Kim PS. Core structure of gp41 from the HIV envelope glycoprotein. Cell. 1997;89(2):263-73. PubMed PMID: 9108481.

12. Chan DC, Chutkowski CT, Kim PS. 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. Epub 1998/12/23. PubMed PMID: 9861018; PubMed Central PMCID: PMC28092.

13. Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, Leunissen JA. Primer3Plus, an enhanced web interface to Primer3. Nucleic Acids Res. 2007;35(Web Server issue):W71-4. doi: 10.1093/nar/gkm306. PubMed PMID: 17485472; PubMed Central PMCID: PMC1933133.

14. Lunter G, Goodson M. Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. Genome research. 2011;21(6):936-9. doi: 10.1101/gr.111120.110. PubMed PMID: 20980556; PubMed Central PMCID: PMC3106326.

15. Studier FW. Protein production by auto-induction in high density shaking cultures. Protein expression and purification. 2005;41(1):207-34. PubMed PMID: 15915565.

16. Johnsson B, Lofas S, Lindquist G. Immobilization of proteins to a carboxymethyldextran-modified gold surface for biospecific interaction analysis in surface plasmon resonance sensors. Anal Biochem. 1991;198(2):268-77. PubMed PMID: 1724720.

17. Otwinowski Z, Minor W. Processing of X-ray Diffraction Data Collected in Oscillation Mode. Methods in Enzymology. 1997;276:307-26.

18. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. Journal of applied crystallography. 2007;40(Pt 4):658-74. doi: 10.1107/S0021889807021206. PubMed PMID: 19461840; PubMed Central PMCID: PMC2483472.

19. Jones TA, Zou JY, Cowan SW, Kjeldgaard M. Improved methods for building protein models in electron density maps and the location of errors in these models. Acta crystallographica Section A, Foundations of crystallography. 1991;47 (Pt 2):110-9. PubMed PMID: 2025413.

20. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta crystallographica Section D, Biological crystallography. 2004;60(Pt 12 Pt 1):2126-32. doi: 10.1107/S0907444904019158. PubMed PMID: 15572765.

21. Afonine PV, Grosse-Kunstleve RW, Adams PD. The Phenix refinement framework. CCP4 Newsletter. 2005;42(Contribution 8).

22. Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, et al. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res. 2007;35(Web Server issue):W375-83. doi: 10.1093/nar/gkm216. PubMed PMID: 17452350; PubMed Central PMCID: PMC1933162.

23. Cashin K, Roche M, Sterjovski J, Ellett A, Gray LR, Cunningham AL, et al. Alternative coreceptor requirements for efficient CCR5- and CXCR4-mediated HIV-1 entry into macrophages. J Virol. 2011;85(20):10699-709. doi: 10.1128/JVI.05510-11. PubMed PMID: 21835796; PubMed Central PMCID: PMC3187472.

24. Eckert DM, Kim PS. Design of potent inhibitors of HIV-1 entry from the gp41 N-peptide region. Proc Natl Acad Sci U S A. 2001;98(20):11187-92. Epub 2001/09/27. doi: 10.1073/pnas.201392898 98/20/11187 [pii]. PubMed PMID: 11572974; PubMed Central PMCID: PMC58705.

25. Weng Y, Weiss CD. Mutational analysis of residues in the coiled-coil domain of human immunodeficiency virus type 1 transmembrane protein gp41. J Virol. 1998;72(12):9676-82. PubMed PMID: 9811701; PubMed Central PMCID: PMC110477.

CHAPTER 5

DESIGN AND CHARACTERIZATION OF FILOVIRUS GP PRE-HAIRPIN INTERMEDIATE MIRROR-IMAGE DRUG TARGETS

Tracy R. Clinton, Matthew T. Weinstock, Michael T. Jacobsen, Nicolas Szabo-Fresnais, Andrew S. Herbert, Laura I. Prugar, Rena McKinnon, Brett D. Welch, John M. Dye, Debra M. Eckert, Michael S. Kay

Introduction

The filoviruses Ebola virus (EboV) and Marburg virus (MarV) are enveloped, negative-strand RNA viruses that cause severe hemorrhagic fever (1). Since their identification almost 50 years ago, there have been over 20 reported natural filovirus outbreaks and several accidental laboratory exposures with an overall mortality rate of 70%, including 4 outbreaks in 2012 (2). Currently no vaccines or therapeutics are FDA approved. Because of ease of transmission, high mortality, and potential for a severe impact on public health, the CDC places filoviruses in its highest category of potential agents of bioterrorism (3). There is a vital need for a broad spectrum filovirus preventative and/or therapeutic to protect against future natural, accidental, or deliberate outbreaks.

Filovirus entry into host cells is mediated by the viral surface glycoprotein (GP), a class I fusion protein. GP is comprised of two disulfide-linked subunits, one surface exposed (GP1) and one embedded in the viral membrane (GP2) (4, 5). Following binding to host cells via cell surface attachment factors, the virus is endocytosed. Endosomal cysteine proteases, such as cathepsins, remove much of GP1 via cleavage, exposing the binding site for the receptor (the endosomally located NPC1) (6-10). At this point, the fusion mechanism is thought to mimic that used by other well-characterized viral class I fusion proteins, such as HIV-1 and influenza (11-13) (Fig. 5.1). GP2 forms a transient conformation ("pre-hairpin intermediate") embedded in both virus (via the transmembrane domain) and host cell (via the fusion loop) membranes. This pre-hairpin intermediate exposes a trimeric coiled coil formed by the N-terminal region (N-trimer) and the unstructured C-terminal region (C-peptide). Slow collapse of the intermediate

into a highly stable trimer-of-hairpins structure, with the C-peptide binding into the grooves on the N-trimer, juxtaposes the virus and cell membranes, leading to membrane fusion.

In the case of HIV-1, the pre-hairpin intermediate has been exploited to develop highly potent viral entry inhibitors (14-16) (Fig. 5.1). Essentially, peptides and proteins that bind with high affinity to either the N-trimer or C-peptide regions prevent formation of the 6-helix bundle in a dominant-negative manner, thereby halting viral entry (as reviewed (17)). The most potent of the HIV entry inhibitors, chol-PIE12-trimer, binds to the highly conserved hydrophobic pockets of the HIV N-trimer and inhibits HIV entry with picomolar potency (14).

Chol-PIE12-trimer is a D-peptide now under preclinical development for HIV treatment and prevention. D-peptides have several important potential advantages as drug candidates (as reviewed (18)). As peptides, they are capable of preventing large protein/protein interactions, something that is generally not possible for small molecules. This gives them the potential of being highly potent (strong clinical efficacy) and specific (low toxicity). In addition, since they are resistant to protease degradation (19), D-peptides should enjoy a long *in vivo* half-life and reduced immunogenicity (20). Specifically for filovirus entry inhibition, resistance to endosomal proteases is highly advantageous.

Since filoviruses and HIV-1 share a similar mechanism of entry, they are likely vulnerable to similar inhibition strategies of targeting the pre-hairpin intermediate. Our ultimate goal is to mimic our HIV-1 entry inhibitor discovery program (14-16, 21) and identify D-peptides that target the filovirus N-trimer and inhibit filovirus entry. The N-

trimer is a promising target as it is highly conserved among filoviruses. Indeed, although the overall sequence identity of GP in the EboV strains pathogenic to humans is only 42%, the N-peptide region is 92% identical (and all changes are conservative). Comparing EboV to MarV, the N-trimer is 67% identical, compared to 26% for GP as a whole. Here, we describe the foundational steps of this development program. Specifically, we have developed and characterized peptide mimics of filovirus N-trimers, validated their use as drug discovery tools, and explored specific conditions that will be applied to our mirror-image phage display (22) discovery process. In addition, using one of our N-trimers as an inhibitor itself (to target the C-peptide region of the pre-hairpin intermediate), we have demonstrated the vulnerability of the pre-hairpin intermediate.

Results and Discussion

N-trimer mimic design

Based on our previous HIV-1 work (21, 23), we set out to design soluble peptide mimics of the N-trimer region of the EboV pre-hairpin intermediate by fusing stable, soluble designed trimeric coiled coils to the N-trimer sequence. As with HIV-1, the filovirus N-trimers aggregate when produced in isolation. Unlike HIV-1, the EboV N-trimer does not possess an obvious "pocket" ideal for drug targeting. Therefore, we were interested in presenting the entire N-trimer groove as well as a smaller, more conserved region of the N-trimer to provide flexibility in targets and drug screening. Our initial designs, in which we fused the coiled coil $IZ_{m(23)}$ N-terminal to N-trimer segments of 29 and 39 amino acids, were highly aggregated as determined by analytical ultracentrifugation (AUC) sedimentation equilibrium experiments (data not shown). To overcome this problem, we fused an additional trimeric coiled coil, GCN4-pI_QI' (IQ) (24)

to the C-terminus of the EboV N-trimer. The resulting peptide, eboIZN39IQ presents the full EboV N-trimer (determined from available 6-helix bundle crystal structures (25, 26)) as a trimeric coiled coil, as shown by circular dichroism (CD) (Fig. 5.2a) and AUC (Table 5.1). eboIZN39IQ is highly stable, as it has almost identical CD spectra at 25 °C, 37 °C, and 50 °C (Table 5.1). The ultimate goal for a filovirus N-trimer mimic is to use it as a target in drug screening, specifically mirror-image phage display (22), to identify inhibitors of filovirus entry. Since these inhibitors will have to bind to the virus in the endosome, all of the biophysical analyses were performed at pH 5.8 to mimic the low pH of the endosome.

To produce a smaller target that presents a discrete, 100% identical region of the N-trimer (across all EboV strains, Fig. 5.3), IZ_m was fused to the N-terminal 21 amino acids of the N-trimer, resulting in eboIZN21. Circular dichroism indicates eboIZN21 is a coiled-coil (Fig. 5.2b), and AUC studies show it is largely trimeric with a slight tendency to form higher order aggregates (Table 5.1). As seen with eboIZN39IQ, eboIZN21 is highly stable, showing almost identical CD signatures at 25 °C, 37 °C, and 50 °C (Table 5.1). We also attempted to produce mimics presenting the middle and C-terminal portions of the N-trimer, but unfortunately, they had difficult solubility problems (data not shown) and were not studied further.

As negative controls for binding studies and drug discovery efforts, we produced mutant N-trimer mimics aimed at abolishing the C-peptide binding site. Specifically, three alanines that are found along the C-peptide binding groove were mutated to aspartate, introducing charges along the groove. The resulting peptides are termed eboIZN39IQ(D3) and eboIZN21(D2). Using CD and AUC, we confirmed that these

mutants maintained the highly stable coiled-coil structure and trimeric nature of their wild type counterparts (Fig. 5.2 and Table 5.1).

C-peptide binding characterization

To validate that eboIZN39IQ presents the native conformation of the N-trimer found in the pre-hairpin intermediate, we characterized the binding of its native ligand as seen in the postfusion 6-helix bundle conformation, the Ebola C-peptide (eboC37). Surface plasmon resonance (SPR) analysis (ProteOn, Bio-Rad) of the full-length interaction of eboC37 binding to eboIZN39IQ shows an affinity constant of 7.5 nM (Fig. 5.4), with no binding to the D3 negative control. This tight binding affinity is the same magnitude as the HIV-1 N-trimer/C-peptide interaction (16), and is strongly indicative of eboIZN39IQ presenting a native N-trimer. A shortened C-peptide (eboC24) missing the 13 N-terminal residues of eboC37 binds to eboIZN39IQ with an affinity constant of ~500 nM and no binding to the D3 negative control (data not shown).

Phage display target validation

The ultimate goal for these N-trimer mimics is to use them as targets in mirror-image phage display to identify D-peptide ligands to serve as drug development leads. To validate them as discovery targets, in a series of phage clone binding assays we analyzed their ability to bind to their native binding partners, eboC37 and eboC24, in the context of phage. Importantly, these experiments were designed not just to verify ligand binding but to define the best conditions for future phage display discovery efforts.

Phage display selections can be conducted in two formats, solid-phase and solution-phase. In solid-phase selections, the target is bound to a solid support (in this case, biotinylated Ebola N-trimer mimic is attached to streptavidin-coated magnetic

beads), and then the phage are incubated with the target on that solid support, typically until equilibrium is reached. Since common phage display libraries are multivalent (multiple copies of the library peptide are expressed on the surface of the phage, due to fusion to multicopy coat proteins), avidity effects increase the apparent binding constant of the library peptides. This avidity-induced affinity boost is beneficial when screening naïve phage libraries, where initial binders typically have low target affinities. In solution-phase, the target and phage are incubated in solution, and avidity effects are minimized. Then, the bound complexes are captured through a brief interaction with a solid support (again, in this case, through a biotinylated target and streptavidin beads). At the same concentration of target, solution-phase is typically a more stringent condition than solid-phase. The stringency of solution-phase is useful when screening consensus libraries, where tight binders must be distinguished from a background of specific binders.

In a solution-phase clonal phage ELISA format carried out at pH 5.8 to mimic the endosomal environment and with biotinylated eboIZN39IQ as the target, both eboC37 and eboC24 clonal phage bound to target significantly over background (both empty beads as well as beads with our negative control, eboIZN39IQ(D3)), validating eboIZN39IQ as a potential phage display target (Fig. 5.5). In this format, both C-peptide clones bound at similar levels, although eboC37 had much greater background binding to both negative controls. Interestingly, a relatively high level of nonspecific phage binding to the eboIZN39IQ target was also observed, as indicated by the M13KE empty phage binding (i.e., no peptide displayed).

Using eboIZN21 as a target in a solid-phase eboC24 clonal phage ELISA, binding at up to two orders of magnitude over background was seen. Also, a very low level of M13KE empty phage binding was observed (Fig. 5.6).

To further investigate the phage background binding to eboIZN39IQ, and to compare the two targets, we analyzed empty M13KE phage binding to both targets under varying conditions (Fig. 5.7). In both solid- and solution-phase formats, M13KE phage binding is significantly higher for eboIZN39IQ beads than blank beads, although the difference is lower for solution phase. For the eboIZN21 target, there is no background binding problem, where binding of M13KE phage is similar to both target and blank beads in solid and solution phase. Under more stringent conditions (solution phase, 100 nM target) the M13KE background binding to eboIZN39IQ is minimized, and an affinity difference for eboC24 binding to eboIZN39IQ vs. eboIZN21 can be seen, indicating the importance of stringency conditions in performing phage display selections. This affinity difference is expected, as the 6-helix bundle structures (25, 26) show the binding site of eboC24 extending past the C-terminus of N21.

Importantly, the first step of a phage display discovery process is to screen a naïve phage display library for binding to the desired target. In such a first selection, where the library diversity only partially samples the large potential sequence space (for example, a naïve 12-mer library has $>10^{15}$ possible sequences (20^n), whereas the typical diversity of the phage display library is 10^{10}), the best binders identified are modest, typically with low to mid-micromolar affinities. Therefore, the selection pressure during the phage panning must also be modest. Standard naïve phage display starting conditions are 10 μM target presented on solid-phase (e.g., (21)). As illustrated in Fig. 5.7, M13KE binding to

eboIZN39IQ is nearly saturated at this condition, and therefore, it would not be possible to identify any binding over background. Under the same conditions, the eboIZN21 background binding is three orders of magnitude lower and equal to blank bead binding, ideal starting conditions for naïve phage display. Therefore, eboIZN21 is an optimized target for our mirror-image phage display discovery process. Additionally, the eboC24-phage can serve as an important positive control to use during naïve phage display to validate the conditions used to capture weak, but specific binders. Importantly, in addition to having ideal behavior in phage display, the N21 region is identical across all EboV strains and highly conserved among filoviruses (95% conserved) (Fig. 5.3).

In addition to serving as a target for identifying D-peptide inhibitors, the filovirus N-trimer mimics (in L) may also be successful targets for antibody phage display and/or as immunogens to elicit a broadly neutralizing antibody that targets this highly conserved region. Such a strategy was successful for identifying a broadly reactive HIV-1 antibody, D5 (27). Finally, they could be a useful cell biological tool, as labeled N-trimers could be used in cell culture experiments to track the appearance of the pre-hairpin intermediate during the viral entry event.

Target synthesis in D

Mirror-image phage display is a clever modification of traditional phage display (21, 22). In traditional phage display, a library of phage, each with a unique peptide displayed on its surface, is screened against a target. The phage links phenotype (target binding) to genotype (the DNA encoding the surface peptide is in the genome). In mirror-image phage display, the screening target is made by standard chemical synthesis with D-amino acids and therefore forms the mirror-image structure of the natural L-target. Phage

display using the D-target is performed, and identified L-peptides that bind the D-target are then chemically synthesized with D-amino acids. By the law of symmetry, these D-peptides bind the natural L-target. Therefore, unlike with traditional phage display, mirror-image phage display targets are limited in size to those that can be chemically synthesized. Importantly, even though the length of eboIZN39IQ is beyond the scope of standard solid-phase peptide synthesis (SPPS) techniques, modern chemoselective ligation techniques (28) allow for its assembly from multiple peptide fragments.

At a length of 48 amino acids, D-eboIZN21 is readily synthesized through standard SPPS techniques. D-eboIZN39IQ, at 101 amino acids, was assembled using native chemical ligation (29) and metal-free desulfurization (30). D-eboIZN39IQ was assembled from three synthetic fragments, whereas D-eboIZN39IQ(D3) was assembled from two synthetic fragments. The final peptide products were confirmed by LCMS.

Vulnerability of the filovirus pre-hairpin intermediate

The key to success in applying our D-peptide inhibitor strategy to filoviruses is the presence of a vulnerable pre-hairpin intermediate during filovirus entry. Exogenous C-peptides derived from the transmembrane subunit of the envelope glycoprotein have been used to validate this vulnerable pre-hairpin intermediate in a variety of viruses (e.g., HIV, SARS, and many paramyxoviruses (31-33)). For filoviruses, an early report showed C-peptide inhibition activity at mM concentrations (34), and a more recent report described modest inhibitory activity (mid-μM) of a C-peptide with an endosomal localization tag (35). Our EboV N-trimer, eboIZN39IQ, provides an additional tool with which to explore the vulnerability of the pre-hairpin intermediate. In support of this

strategy, the peptide mimics of the HIV-1 N-trimer inhibit HIV entry at mid-nM concentrations by binding to the C-peptide region of the exposed intermediate.

Indeed, eboIZN39IQ inhibits entry in our pseudovirus system in which Ebola GP is expressed on the surface of an HIV particle, with an IC_{50} of 320 nM (averaged over four replicate experiments) (Fig. 5.8a). Importantly, the anti-EboV activity of our negative control, eboIZN39IQ(D3), is >30-fold diminished at an IC_{50} of 11 uM (averaged over three replicate experiments). The residual activity of eboIZN39IQ(D3) is unlikely to be due to specific pre-hairpin intermediate binding activity, since SPR analysis showed essentially no binding of eboIZN39IQ(D3) to the C-peptide (Fig. 5.4). However, it is difficult to determine the exact nature of the modest eboIZN39IQ(D3) activity, as it is not seen against VSV, and no morphological changes were observed via light microscopy in any of the wells throughout the assay. eboIZN39IQ demonstrates modest activity against MarV, at an IC_{50} of 5.7 μM (averaged over duplicate experiments); however, it is only ~2-fold better than the eboIZN39IQ(D3) anti-MarV activity. It is therefore difficult to determine if this indicates any specific activity due to cross reactivity with the MarV pre-hairpin intermediate.

The ability of eboIZN39IQ to inhibit the entry of wild-type EboV and MarV was also assessed using a filovirus immunofluorescence assay under BSL4 conditions (Fig. 5.8b). Although eboIZN39IQ is significantly less potent in this assay, there is good inhibition of entry at the highest concentration tested, 10 μM, and no inhibition by our negative D3 control. Also, no activity was seen against authentic MarV. Taken together, these data validate the presence of a vulnerable pre-hairpin intermediate during the filovirus entry process.

Unlike HIV-1, filoviruses enter cells via endocytosis and initiate membrane fusion late in the endosomal process. Therefore, filovirus entry inhibitors will have to enter into and be active in the endosomes. Although eboIZN39IQ does not possess a specific tag to localize it to endosomes, it is highly charged on its surface and may naturally interact with the cell membrane, allowing it to access the endosome more efficiently than C-peptides. Interestingly, the inhibitory activity we observe in both the pseudovirus and authenic EboV systems is dependent on the presence of DEAE-dextran. This cationic polymer may reduce electrostatic repulsion between eboIZN39IQ and the negatively charged cell membrane surface. As a highly structured peptide, it would also likely survive against proteolysis longer than unstructured C-peptides.

Endosomal targeting

In order to achieve high potency, filovirus inhibitors will likely need to be targeted to the endosome. A variety of cationic tags have been described to traffic to or through the endosomes, such as HIV-1 Tat and poly-arginine sequences (as reviewed (36)), and Tat was required to confer inhibitory activity on an EboV C-peptide (35). However, we have tried a variety of these cationic tags, including one optimized for endosomal localization (37), on both EboV C-peptides and N-trimer mimics and none of them have convincingly improved inhibitory activity in a specific way (unpublished results). Also, each of them contributed to cellular toxicity at concentrations slightly higher than observed inhibitor IC_{50}s, visualized as morphological differences under the light microscope and/or slowed growth.

As an alternate strategy, we have fused cholesterol to our N-trimer mimic, and it appears to increase inhibitor potency in a specific fashion (preliminary data not shown).

Indeed, cholesterol conjugation increased the potency of our HIV-1-specific PIE12-trimer by two orders of magnitude by targeting it to cellular membranes (14). Further work needs to be done to confirm these preliminary data and to determine if this strategy will be useful for future inhibitors, especially in light of a recent paper that describes non-specific inhibitory activity for a cholesterol-conjugated filovirus C-peptide (38).

In the future, we will likely focus on targeting inhibitors to endosomes in a more specific manner. For example, transferrin receptor (TfR) transports iron into cells via an endosomal pathway. TfR is ubiquitously expressed, and is frequently used as an endosomal marker. In addition, many cancer therapies that use the TfR for targeted delivery are in development (as TfR is overexpressed in many cancer cells) (39). Specifically, a 7-residue peptide was identified via phage display to bind to TfR and has been used to target cancer therapies in preclinical studies (39-41). Fusing this peptide to a D-peptide that targets the filovirus pre-hairpin intermediate is one potential strategy for targeting inhibitors to the endosome.

Conclusion

We have developed an ideal filovirus N-trimer target to be used in mirror-image phage display, eboIZN21. It is a stable trimeric coiled coil with low background phage binding, and easily synthesized in D. With eboIZN21 and a good positive control phage clone (eboC24-phage), we are poised for phage display selections.

Materials and Methods

Reagents

Plasmids and cells were obtained from the indicated sources: pKA8 vector (gift from C. Hill), pEBB-HXB2 (gift from B. Chen), SV-ZeboGPΔmuc and SVMarVGP (gift from M. Farzan) (42), BLR(DE3)pLysS *E. coli* (Novagen), BL21-Gold(DE3)pLysS *E. coli* and XL-1 Blue *E. coli* (Stratagene). pNL4–3.Luc.R-E- (N. Landau) (43, 44) and HOS-CD4-fusin (N.Landau) (45, 46) were obtained from the NIH AIDS Research and Reference Program. The mammalian cells were propagated in standard tissue culture medium, Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% fetal calf serum (Invitrogen).

Recombinant peptide production and purification

eboIZN39IQ and eboIZN39IQ(D3) were produced via PCR gene synthesis. The IZ_m and IQ fragments were PCR amplified from Kay lab plasmids encoding HIV-1 N-trimer mimics. An NdeI site was included in the 5' PCR primer for IZ_m, and a BamHI site was included in the 3' PCR primer for IQ. The EboV N39 sequence was synthesized in two overlapping oligos and companion primers. All internal primers contained complementary sequences so the three separate components, IZ_m, N39, and IQ could be annealed and amplified together. The resulting DNA fragment was cloned into the NdeI/BamHI cloning sites of pKA8 and expressed in BLR(DE3)pLysS cells using an autoinduction protocol. Specifically, cultures were inoculated from a single colony and grown overnight at 37 °C in autoinduction media (47). The resultant peptide has an N-terminal His tag (His_8) followed by a TEV cleavage site (ENLYFQG). A single tyrosine was placed at the end of the sequence to facilitate concentration determination via

absorbance at 280 nm. The peptides were resuspended from inclusion bodies using a Ni++ Binding Buffer (20 nM Sodium Phosphate pH 8.0, 300 mM NaCl, 10 mM imidazole) + 6 M GuHCl, and purified via gravity flow Ni++ affinity chromatography (HIS-Select Nickel Affinity Gel, Sigma Aldrich). The purified peptide was dialyzed into 5% acetic acid and further purified by reverse phase HPLC on a C18 column (Vydac) and lyophilized. Peptide powder was resuspended in water and then diluted to 0.2 mg/mL in 50 mM Sodium Phosphate pH 6.5, 0.5 mM EDTA, 1 mM DTT and digested with TEVse (gift of C. Hill) overnight at 30 °C. The digested peptide was dialyzed into 5% acetic acid and then HPLC purified and lyophilized. The final peptide sequences are: GHMD**IKKEIEAIKKEQEAIKKKIEAIEKE***LRQLANETTQ(A/D)LQLFLR(A/D)TTEL RTFSILNRK(A/D)IDFLLQR***MKQIEDKIEEIESKQKKIENEIARIKKLIGER**Y, with IZ_m and IQ shown in bold, the EboV N-trimer in italics, and the three alanine positions that are changed to aspartate in the D3 mutant in parentheses.

Biotinylated eboIZN39IQ and eboIZN39IQ(D3) for SPR analysis and phage display were expressed from plasmids that are slightly modified to those described above. Using PCR, a CGG sequence was added N-terminal to IZ (GHMCGGDIKK...). Expression and purification was as described above with additional reduction steps included to keep the cysteine reduced during purification. The purified protein was biotinylated with EZ-link Maleimide-PEG2-biotin (Thermo Scientific). The purified lyophilized powder was resuspended at 1 mM in freshly prepared reaction buffer (6 M GuHCl, 150 mM NaCl, 100 mM Na_2HPO_4, 5 mM TCEP) and the biotinylation reagent was added at 5 mM and allowed to react for 4 hr at RT. The biotinylated peptides were

purified by reverse phase HPLC on a C18 column (Waters) and lyophilized. The mass of the peptide was confirmed by LCMS (AB Sciex API 3000 LC/MS/MS system).

Peptide synthesis

eboIZN21, eboIZN21(D2), eboC37, and eboC24 were chemically synthesized using solid-phase peptide synthesis (SPPS) with Fmoc-amino acids (CBL Biopharma) on the Prelude peptide synthesizer (PTI). They were synthesized on TentaGel R RAM resin (Rapp Polymere) to yield C-terminal amide-capped peptides. Standard synthesis scales were 25 - 32 μmol per peptide. Standard amino acid coupling was as follows: 3x3 min deprotection with 20% piperidine in DMF followed by 25 min couplings with 72.2 mM amino acid (200 mM stocks in NMP), 71.5 mM HATU (200 mM stock in DMF), and 166.7 mM NMM (600 mM stock in DMF). Biotinylation with N-Biotinyl-NH-$(PEG)_2$-COOH DIPEA (20 atoms) (Novabiochem) was coupled for 2 hrs with identical reagent composition to a standard coupling. N-terminal capping was accomplished in 30 min with 2 mL acetic anhydride and 2 mL 0.6 M NMM either on- or off-line. Peptide cleavage from resin was accomplished off-line with 92.5% TFA, 2.5% EDT, 2.5% TIS, 2.5% H_2O when the peptide contained Met or Cys residue(s) or with 95% TFA, 2.5% TIS, 2.5% H_2O in the absence of any Met/Cys residues followed by precipitation/washing with diethylether. All peptides were purified by reverse-phase HPLC on a Waters BEH X-Bridge C18 column (10 μm, 300 Å, 19 x 250 mm) with a water/ACN gradient in 0.1% TFA. All peptides were lyophilized and their molecular weight verified by LCMS (AB Sciex API 3000 LC/MS/MS system).

D-eboIZN39IQ was assembled from three synthetic peptide fragments via native chemical ligation/metal-free desulfurization. Peptides were synthesized via Fmoc-SPPS

on a PTI PS3 peptide synthesizer at 100 umol scale. The C-terminal peptide was synthesized on Rink Amide AM resin LL (Novabiochem) and the other two fragments were synthesized on Dawson Dbz AM resin (Novabiochem). The C-terminal fragment contained an N-terminal Fmoc-Cysteine-OH residue in the place of a native alanine for native chemical ligation (CIDFLLQRMKQIEDKIEEIESKQKKIENEIARIKKLIGERY). For the same reason, the middle fragment contained an N-terminal Boc-L-thiazolidine-4-carboxylic acid (Boc-THZ-OH, Bachem) as its N-terminal residue in place of the native alanine at that position ((THZ)-NETTQALQLFLRATTELRTFSILNRK). The N-terminus of the N-terminal peptide (GHMDIKKEIEAIKKEQEAIKKKIEAIEKELRQL) was biotinylated with N-Biotinyl-NH-(PEG)2-COOH DIPEA (Novabiochem). For peptides synthesized on Dawson Dbz AM resin, the C-terminal linker was converted to the resin bound benzimidazolinone (Nbz) according to manufacturer recommendations. Cleavage of all peptides was performed according to standard procedures. Peptides were purified by reverse-phase HPLC on a Waters BEH X-Bridge C18 column (10 μm, 300 Å, 19 x 250 mm) with a water/acetonitrile gradient in 0.1% TFA. Ligations were performed according to (48). Following the ligation between the C-terminal and middle fragments, the N-terminal THZ residue was converted to cysteine by dissolving the purified ligation product in 6 M GuHCl, 200 mM sodium phosphate, 200 mM methoxyamine HCl, pH 4. After THZ to Cys conversion was achieved, the buffer was brought to 200 mM MPAA and 20 mM TCEP, the pH was adjusted to 7, and the N-terminal peptide was added to the solution for the final ligation. Following purification of the ligation product by reverse-phase HPLC, the cysteine residues at the ligation junctions were converted to the native alanine residues via a metal-free, radical-mediated desulfurization strategy essentially as

described in (30) except *t*BuSH was replaced with glutathione and desulfurizations were performed at 37ºC. eboIZN39IQ(D3) was synthesized in an analogous (though simplified) manner from two peptide fragments.

Preparation of peptide samples for biophysical analysis

For biophysical analyses, peptide stocks were prepared in water from lyophilized peptide at concentrations of 400 μM or greater for a minimum absorbance at 280 nm of 0.1 in a 1 cm pathlength cuvette. Stocks were centrifuged at 18,000 xg for 10 min to remove aggregates. Absorbance at 280 nm (assuming ε_{280} of 1408 $M^{-1}cm^{-1}$ for tyrosine) was used to determine stock concentrations (49). For eboIZN39IQ and eboIZN39IQ(D3), both recombinant and synthetic, UV absorbance consistently overestimated the concentration of the stocks (as evidenced by an unusually high 260/280 ratio as well as CD traces whose shape depicted ideal coiled coils but whose signal had a lower than expected absolute value). Therefore, the concentrations of these stocks were determined via quantitative amino acid analysis (D. Winge). The peptides were then diluted to the desired concentration in 50 mM Sodium Phosphate pH 5.8, 150 mM NaCl. For eboIZN21 and eboIZN21(D2), all experiments described in this paper were performed with biotinylated peptide. For eboIZN39IQ and eboIZN39IQ(D3), CD, AUC, and viral infectivity were performed with nonbiotinylated material, whereas SPR and phage display used biotinylated material.

Circular dichrosim

Circular dichrosim (CD) data were obtained using an AVIV Model 410 spectrophotometer (AVIV, Lakewood, NJ). Samples were analyzed in a 1 mm pathlength

quartz cuvette at 25, 37, and 50 °C. Prior to CD analysis, prepared samples (in 50 mM Sodium Phosphate pH 5.8, 150 mM NaCl) were centrifuged at 18,000 Xg for 10 min to remove aggregates. CD data were scanned in triplicate and buffer subtracted. Final CD data were presented according to mean residue ellipticity equation, $[\theta] = 100*\theta/[(n-1)*(\ell)*(c)]$, where θ is observed ellipticity, n-1 is number of peptide bonds, ℓ is the pathlength in cm, and c is the peptide concentration in mM. Due to aggregation observed with eboIZN39IQ and eboIZN39IQ(D3) upon initial dilution in CD buffer, their final concentrations were corrected from the original amino acid analysis values based on the ratio of ellipticity at 222 nm post- and precentrifugation ($\theta_{222\text{-post-spun}}/\theta_{222\text{-pre-spun}}$).

Analytical ultracentrifugation sedimentation equilibrium

Sedimentation equilibrium analysis was performed on each peptide at three concentrations each (a starting concentration and two 2-fold dilutions, with typical starting concentrations between 10-30 μM). Dilutions were prepared in matching buffer (50 mM Sodium phosphate, 150 mM NaCl, pH 5.8), and the same buffer was used for blanks. Each sample was spun until equilibrium, typically ~24 hrs, at a minimum of two speeds, but typically three speeds (18,000, 21,000, and 24,000 RPM). Each data set was globally fit to a single ideal species with a nonlinear least squares algorithm as implemented in HETEROANALYSIS (50). Buffer densities and protein partial specific volumes were calculated with SEDNTERP (version 1.09) (51). For the biotinylated peptides, partial specific volumes were adjusted based on reported values for PEG (52).

Surface plasmon resonance

SPR analysis was conducted on the Bio-Rad ProteOn instrument in PBS* running buffer (50 mM Sodium Phosphate, 150 mM NaCl, pH 5.8) + 0.1 mg/mL BSA and 0.01% Tween-20. Approximately 600 RUs of Biotin-eboIZN39IQ and Biotin-eboIZN39IQ(D3) targets (200 nM stocks ultracentrifuged for 30 min at 45,000 rpm) were loaded at 66 - 200 nM onto the NLC neutravidin-coated chip. Using the one-shot kinetics method, a 2-fold dilution series was performed in triplicate at RT starting at 60 nM for eboC37 and a 3-fold dilution series in triplicate at RT starting at 5 μM for eboC24. 10 min dissociation time for eboC37 and 5 min dissociate time for eboC24 was used to ensure the response fully recovered to baseline prior to the next injection. Data were fitted to the Langmuir binding model using ProteOn Manager software (Bio-Rad).

Clonal phage production

Forward and reverse sandwich oligos encoding the C-peptide clones were designed based on the primary sequence of each clone. The oligo sandwich was annealed with 5 μg of each primer in 50 μL total volume in ddH_2O by heating to 95°C and slow cooling and then extended with Klenow Fragment (NEB). The inserts and M13KE cloning vector backbone (NEB) were digested with Acc65I and EagI-HF. The insert DNA was EtOH precipitated and then gel purified from a 6% TBE acrylamide gel using a Bio-Rad Freeze n Squeeze device. The inserts and backbone were ligated and transformed into SS320 electrocompetent cells. After brief recovery, the transformed cells were added to fresh XL-1 Blue cells and plated on LB/IPTG/X-gal plates. The DNA from specific phage plaques was PCR amplified and Sanger sequenced (Eton

Biosciences), and those containing the correct DNA were subsequently amplified from a single plaque.

Phage amplification

A single plaque was added to XL-1 Blue cells (OD_{600} 0.5 – 1), diluted to 40 mL of OD_{600} 0.05 in LB + 25 μg/mL tetracycline, and shaken at 220 RPM at 37°C for 4.5 – 5 hrs. Cells were pelleted by centrifugation, and the phage supernatant was sterile filtered. Phage were precipitated with PEG-NaCl (20% w/v polyethylene glycol-8000 (Fisher Scientific), 2.5 M NaCl) overnight at 4°C. Precipitated phage were then pelleted via centrifugation and resuspended in TBS (50 mM Tris-HCl, 150 mM NaCl, pH 7.4). They were PEG-precipitated again (~1 hr on ice), centrifuged, and resuspended in 200 μL TBS. Aliquots were flash frozen and stored at -20°C with a working stock left at 4°C if imminent experiments were planned.

Clonal phage ELISA

For each phage ELISA reaction, 30 μL Streptavidin-coated magnetic beads (Invitrogen, 10 mg/mL Dynabeads MyOne, Streptavidin T1) at 10 mg/mL were magnetically pelleted and washed with 3.33x bead volume TBS. The beads were then blocked in 3.33x bead volume 100% SB (ThermoScientific, SuperBlock Blocking Buffer in TBS, pH 7.4) for 10 min at RT and rinsed with equal volume of 100% SB* (SB adjusted to pH 5.8 with HCl). Solution-phase beads were then resuspended in 3.33x bead volume of 100% SB* and stored at 4°C for up to 24 hrs. Solid-phase beads were resuspended in 3.33x bead volume PBS* (50 mM Sodium Phosphate, 150 mM NaCl, pH 5.8) + 10% SB*. To load target onto solid-phase beads, 1x bead volume of an appropriate

target concentration (e.g., 10 μM target for 10 μM beads) was added and incubated for 10 min followed by adding 3.33x bead volume 5 mM D-Biotin (in PBS* + 10% SB*) and incubating for an additional 5 min. For blank (no target) beads, 3.33x bead volume of 5 mM D-Biotin was added and incubated for 5 min. All beads were then magnetically pelleted, washed in PBS* and resuspended in 1x bead volume PBS* + 10% SB*.

Solid-phase binding reactions were incubated in 96-well format (Costar, sterile polystyrene, V-bottom, nontreated) with shaking for at least 2 hrs at RT either as 30 or 100 μL reactions in 1x PBST* (50 mM Sodium phosphate, 150 mM NaCl, pH 5.8, 0.01% Tween-20) + 10% SB* and 10^{10} phage clone added. All washes and elution were done on the KingFisher Duo magnetic particle processor (Thermo Scientific). The binding reaction was mixed on the KingFisher for 1 min at medium speed and the beads collected by 5 sec dips of the magnet through the sample, repeated 5 times (5 x 5 sec). All washes were done with PBST* (wash 1: 700 μL; wash 2: 800 μL; wash 3: 900 μL; washes 4-7: 1000 μL), mixed at slow speed for 1.5 min, and beads collected 3 x 3 sec. Bound phage were eluted with 50 μL EB (0.2 M glycine, pH 2.2) for 10 min, beads collected 5 x 5 sec, and neutralized with 7.5 μL NB (1 M Tris, pH 9.1). Dilutions of eluted phage were used to infect XL-1 Blue cells and then plated in top agar (40% LB agar/60% LB) on LB/IPTG/X-gal plates (LB agar, 25 μg/mL tetracycline, 1 mM IPTG, 0.1 mg/mL Xgal). Blue plaques were then counted to determine phage titers.

Solution-phase binding reactions were performed similarly to solid-phase 30 μL reactions. Instead of adding target-loaded beads to the binding reaction, an appropriate amount of soluble target (10x target added to 10% of total reaction volume) was added to the reaction just before phage were added. Additionally, on the Kingfisher Duo, target

and bound phage were pulled down in a rapid 1 min magnetic pelleting step (1 min slow mixing, 5 x 5 sec bead collect). All washes were done with PBST* except wash 1 which contained 5 mM D-Biotin to block unoccupied streptavidin sites (wash 1: 150 μL; wash 2: 700 μL; wash 3: 800 μL; wash 4: 900 μL; wash 5: 1000 μL), mixed at slow speed for 25 sec, and beads collected 3 x 3 sec.

Pseudovirus infectivity assays

Single-cycle pseudovirions were produced with a pNL4-3 HIV-1 genome with firefly luciferase inserted into the nef gene and frameshift mutations in both Env and Vpr and expressing filovirus GP on their surface (or VSV for a specificity control). These pseudovirions were produced by co-transfecting 293T human embryonic kidney cells with the described HIV-1 genome (pNL4-3.Luc.R-E-) and a plasmid encoding the desired virus glycoprotein (SV-ZeboGPΔmuc for EboV Zaire GP lacking the mucin domain, SV-MarVGP for MarV Musoke GP, and pMDG VSV-G for VSV) in the presence of PEI transfection reagent. Pseudovirus-containing supernatant was collected and filtered 38-43 hrs posttransfection. For EboV and MarV, pseudovirions were concentrated by centrifuging through a 20% sucrose/TNE (10 mM Tris pH 7.6, 100 mM NaCl, 1 mM EDTA) cushion (26,000 RPM, 2 hrs), and then the pellet was resuspended in TNE, aliquoted, and stored at -80 °C.

To measure inhibition of infectivity, 90 uL of each inhibitor dilution and 8.9 ug/mL DEAE-dextran were added to HOS-CD4-fusion cells in a 96-well format. The plates were then transferred to BSL3, and 10 uL of pseudovirus diluted in media was added to each well 30-60 min after the inhibitor addition (final DEAE-dextran concentration of 8 ug/mL). 24 hrs later, all wells were inspected under light microscope

to check for gross morphological changes. Virus and inhibitor were then removed via aspiration, and fresh media was replenished. 20-24 hrs later, the cells were lysed, and the luciferase activity was measured (Bright-Glo luciferase assay system, Promega). To determine IC_{50} values, the data from each inhibitor concentration series were normalized to the noninhibitor control signal and then was fit to a Langmuir equation: $y = 1/(1+[\text{inhibitor}]/IC_{50})$ (Kaleidagraph, Synergy Software). The curve fit was weighted by the normalized standard error of each concentration point (with a minimum error allowed of 1%).

Filovirus immunofluorescence assays

Vero cells were seeded in 96 well black plates. Peptides and vehicle control were diluted to 1.1X final concentration in culture media and incubated on the plate for 1 hr at 37 °C. In the BSL4, 10 uL of virus diluted in media and DEAE-dextran was added to each well (final DEAE-dextran concentration of 8 μg/mL). For EboV Zaire, infections were performed at 1, 0.1, and 0.01 MOI of virus. For MarV, infections were performed at 1.0 MOI only. After 1 hr at 37 °C, the virus and peptides were removed, the wells were washed with PBS, and media with peptide was used to replenish the wells. At 24 hrs postinfection, each well wash visualized via light microscope to look for any gross morphological abnormalities. At 48 hrs postinfection, the wells were washed with PBS and then the cells fixed with 10% formalin. After blocking, the fixed cells were incubated with GP-specific mAb (9g4 for MarV, KZ52 for EboV) followed by incubation with FITC-labeled secondary antibody (goat anti-mouse or anti-human, respectively). Nuclei were stained with Hoechst solution. Cells were imaged using an Operetta high content

device and images were analyzed using Harmony software to determine percent of infected cells in a given well. Data were plotted normalized to the vehicle control.

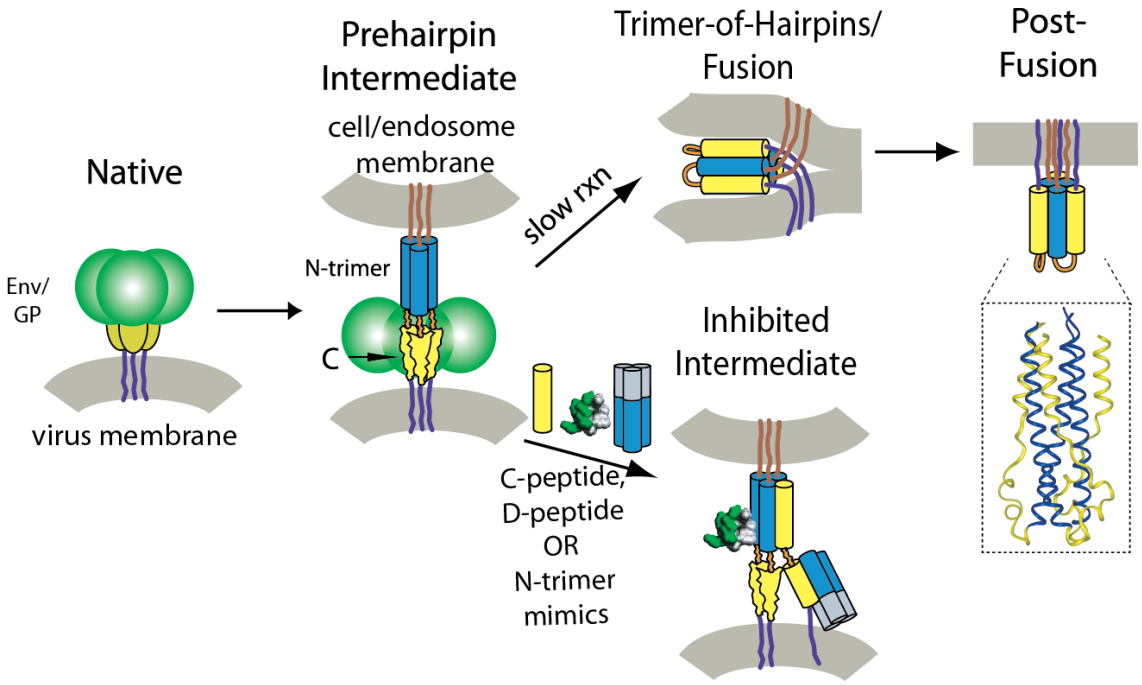


Figure 5.1. Model for membrane fusion mediated by enveloped viral surface glycoproteins. The HIV-1 and filovirus entry events are predicted to be highly similar. First, the surface glycoprotein (Env for HIV-1, GP for filoviruses) facilitates viral attachment to the cell and, for filoviruses, the virus is endocytosed and then cleaved by endosomal proteases. Engagement of the virus receptors (CD4 and a chemokine receptor for HIV-1, NPC1 for filoviruses) leads to a conformational change in Env/GP, and insertion of the fusion peptide/loop into the host cell membrane. At this state, the virus is in a transient state that bridges both membranes, termed the "pre-hairpin intermediate." It is now vulnerable to inhibitors that bind the pre-hairpin intermediate and inhibit entry. In the absence of an inhibitor, the Env/GP structure slowly resolves into the highly stable trimer-of-hairpins structure, juxtaposing the two membranes and leading to membrane fusion. The inset shows the high resolution structure of the EboV trimer-of-hairpins (25). In the HIV model system, it has been shown that inhibitors that bind to either the N-trimer or C-peptide region are capable of inhibiting entry.

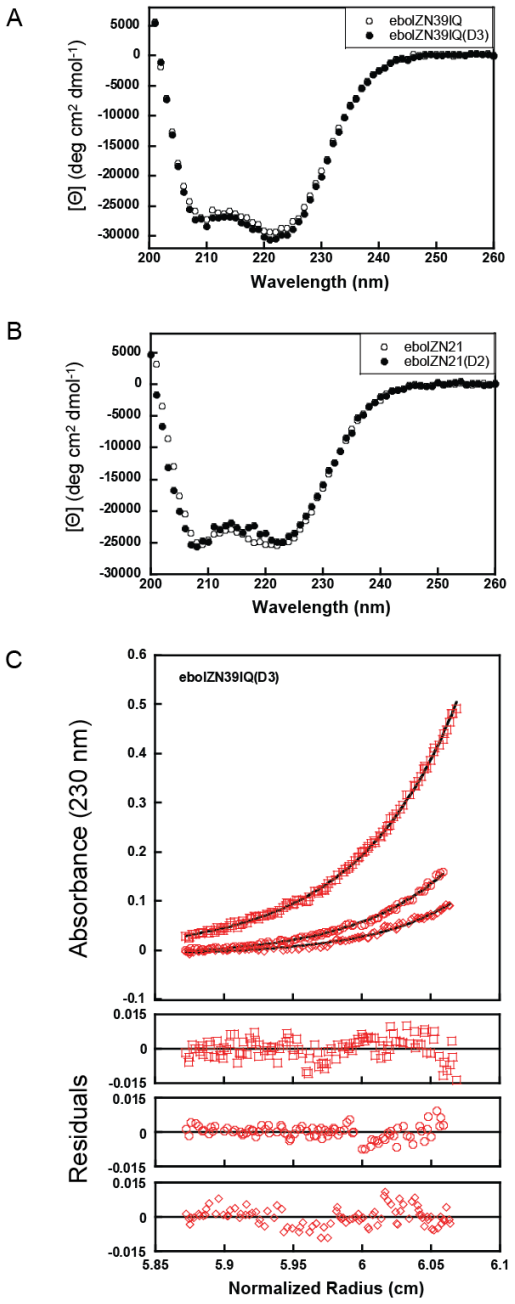


Figure 5.2. Biophysical analyses of EboV N-trimer mimics. A) CD spectra of 11.4 μM eboIZN39IQ and 11.1 μM eboIZN39IQ(D3) at 25 °C. Both spectra indicate a highly helical conformation. B) CD spectra of 18.0 μM eboIZN21 and 25.3 μM eboIZN21(D2) at 25 °C, also indicative of a completely helical conformation. C) Analytical ultracentrifugation (AUC) sedimentation equilibrium analysis of eboIZN39IQ(D3), shown as representative AUC data. 8, 4, and 2 μM peptide was centrifuged at 18,000, 21,000, and 24,000 RPM at 25 °C on a Beckman XLA. Data were globally fit to a single ideal species, and an observed molecular weight of 37,278 Da was determined for an M_{obs}/M_{calc} of 3.07. The data (open symbols) and fit (solid lines) are shown for the highest speed.

```
Zaire        LRQLANETTQALQLFLRATTE
Taï Forest   LRQLANETTQALQLFLRATTE
Bundibugyo   LRQLANETTQALQLFLRATTE
Sudan        LRQLANETTQALQLFLRATTE
Reston*      LRQLANETTQALQLFLRATTE
Marburg      LRRLANQTAKSLELLLRVTTE
Ravn         LRRLANQTAKSLELLLRVTTE
Lloviu*      LRELANTTTKALQLFLRATTE
```

Figure 5.3. Alignment of N21 from GP of the 5 known EboV species, Marburg, Ravn, and Lloviu hemorrhagic fever filoviruses. Conserved residues (score of 0 or higher in BLOSUM62 matrix (53)) highlighted in gray. All changes are conserved except for the E/Q to T substitution in Lloviu. *Reston and likely Lloviu are not pathogenic to humans.

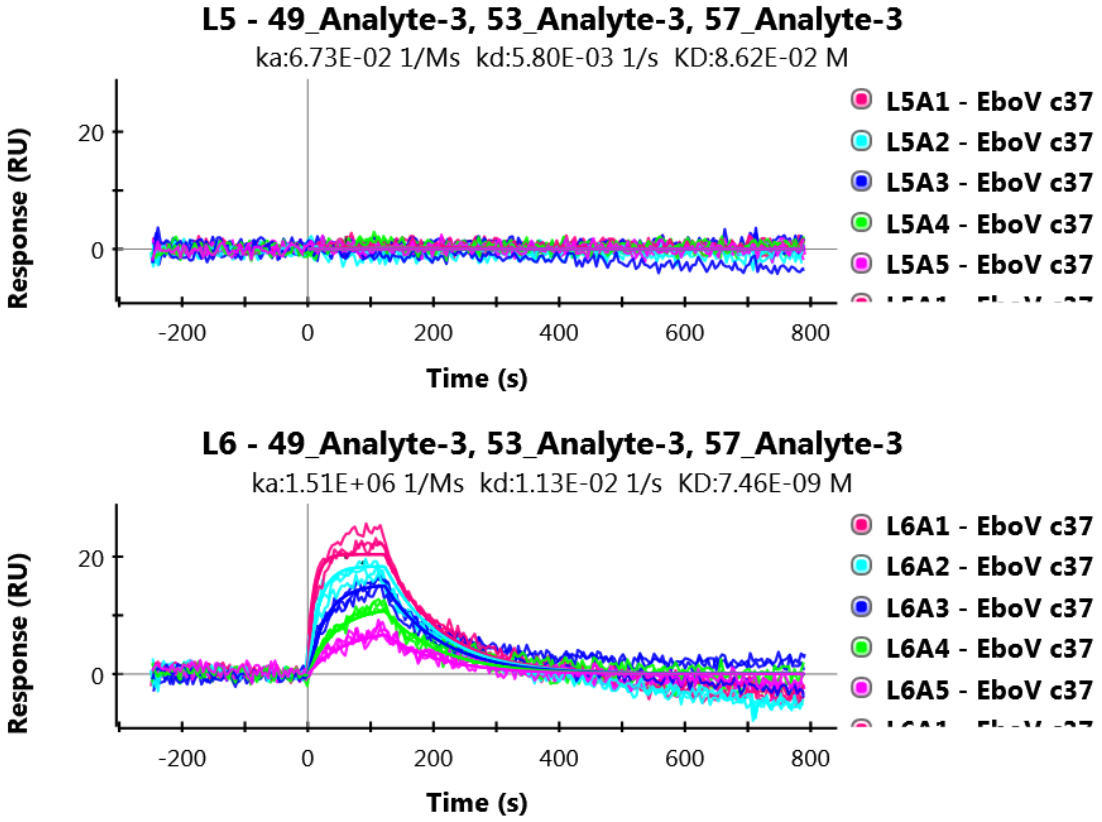


Figure 5.4. Affinity measurement of the EboV C-peptide for the N-trimer mimic. Top: Sensogram of eboC37 flowed over eboIZN39IQ(D3) in a 2-fold dilution series starting at 60 nM. Little to no binding is observed. Bottom: The same eboC37 dilutions flowed over an eboIZN39IQ surface. The fit indicates a K_D of 7.5 nM.

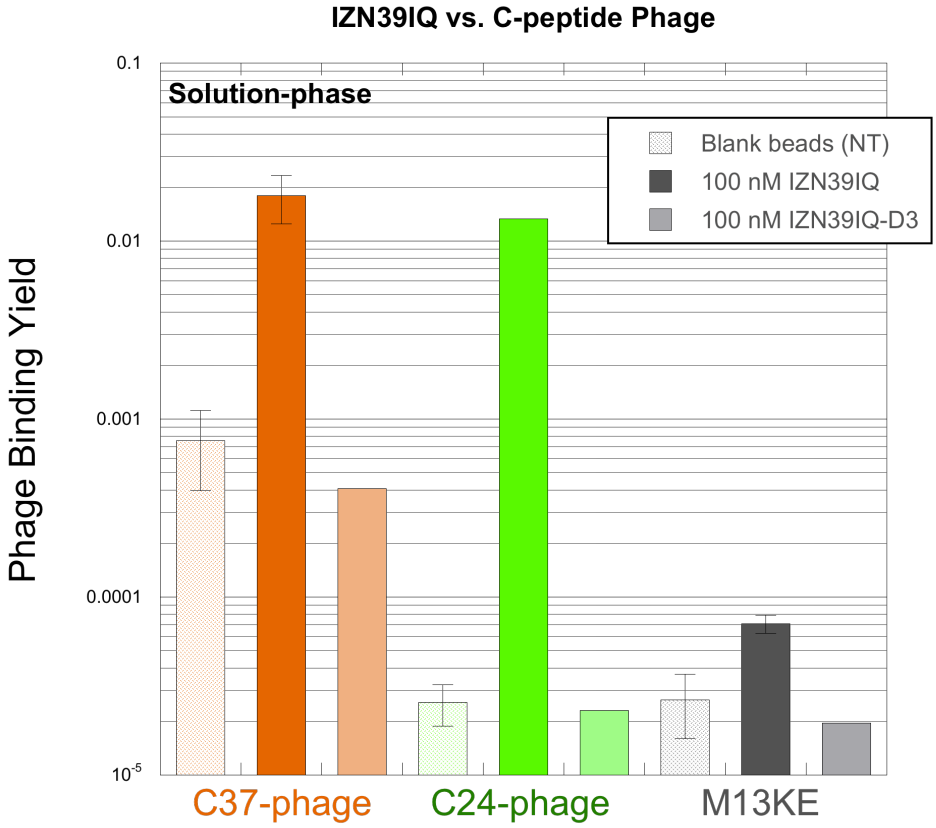


Figure 5.5. Solution-phase clonal phage ELISA of EboV C-peptides binding to eboIZN39IQ. eboC37 (orange) and eboC24 (green) phage binding to blank beads (NT), eboIZN39IQ, and eboIZN39IQ(D3) is shown. Control phage, M13KE, binding was also assessed (gray). Where present, error bars represent standard error across multiple experiments for the C37 and M13KE data (n=3) and standard deviation for the C24 data (n=2).

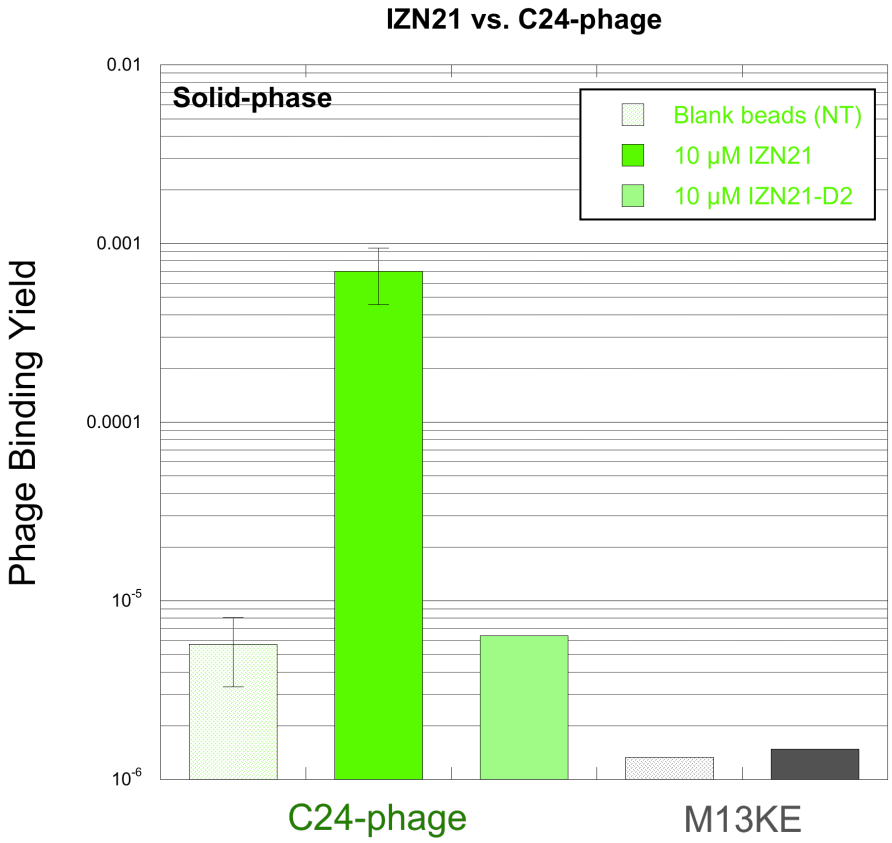


Figure 5.6. Solid-phase clonal phage ELISA validating the eboIZN21 target in the context of phage display. eboC24 (green) phage binding to blank beads (NT), eboIZN21, and eboIZN21(D2) is shown. Control phage, M13KE, binding to blank beads (NT), and eboIZN21 was also determined. Where indicated, error bars represent the standard error of data from four experiments.

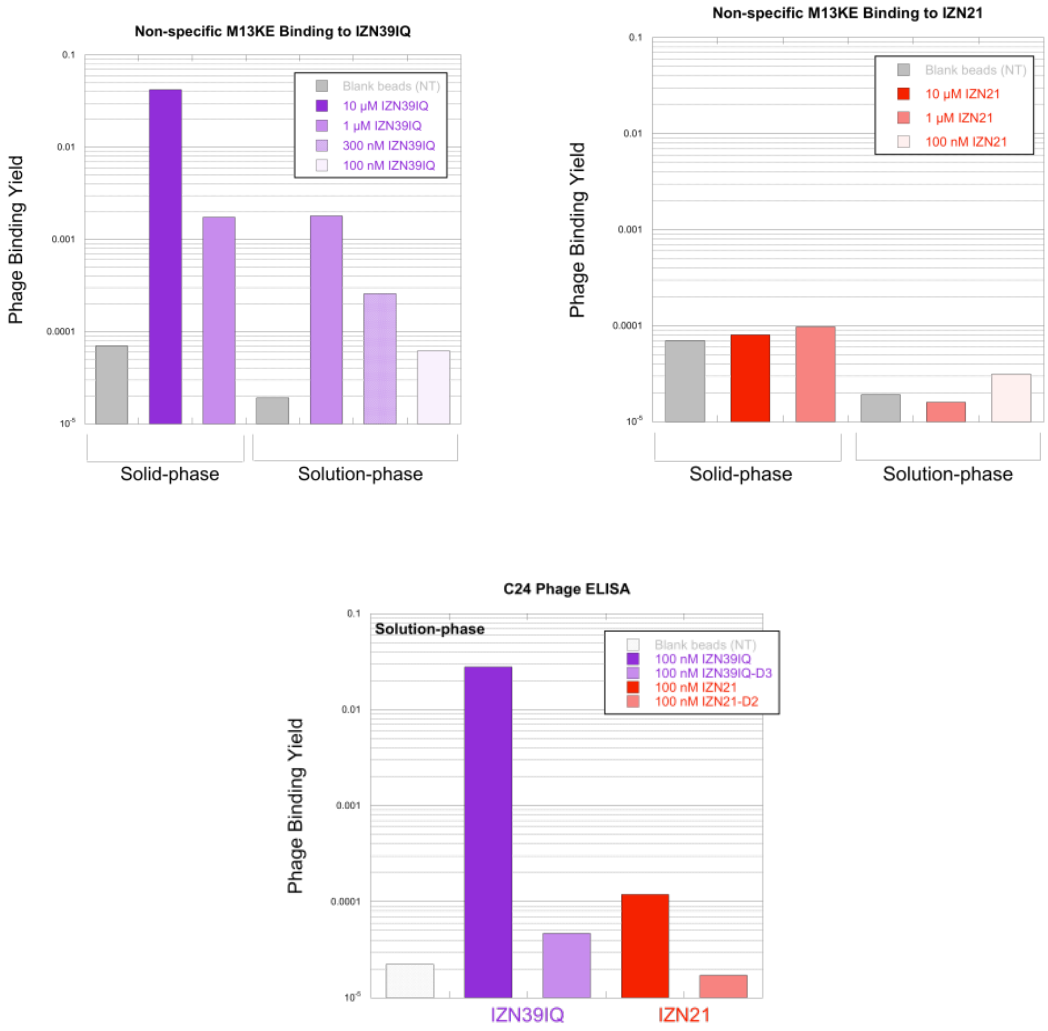


Figure 5.7. Comparing the two EboV N-trimer mimics as phage display targets. Phage background binding is greater to eboIZN39IQ than to eboIZN21. Phage ELISA showing M13KE control phage binding to eboIZN39IQ (left) and eboIZN21 (right) under both solid-phase and solution-phase conditions. High stringency solution-phase binding shows an affinity difference for eboC24. Experiments were repeated and representative data shown.

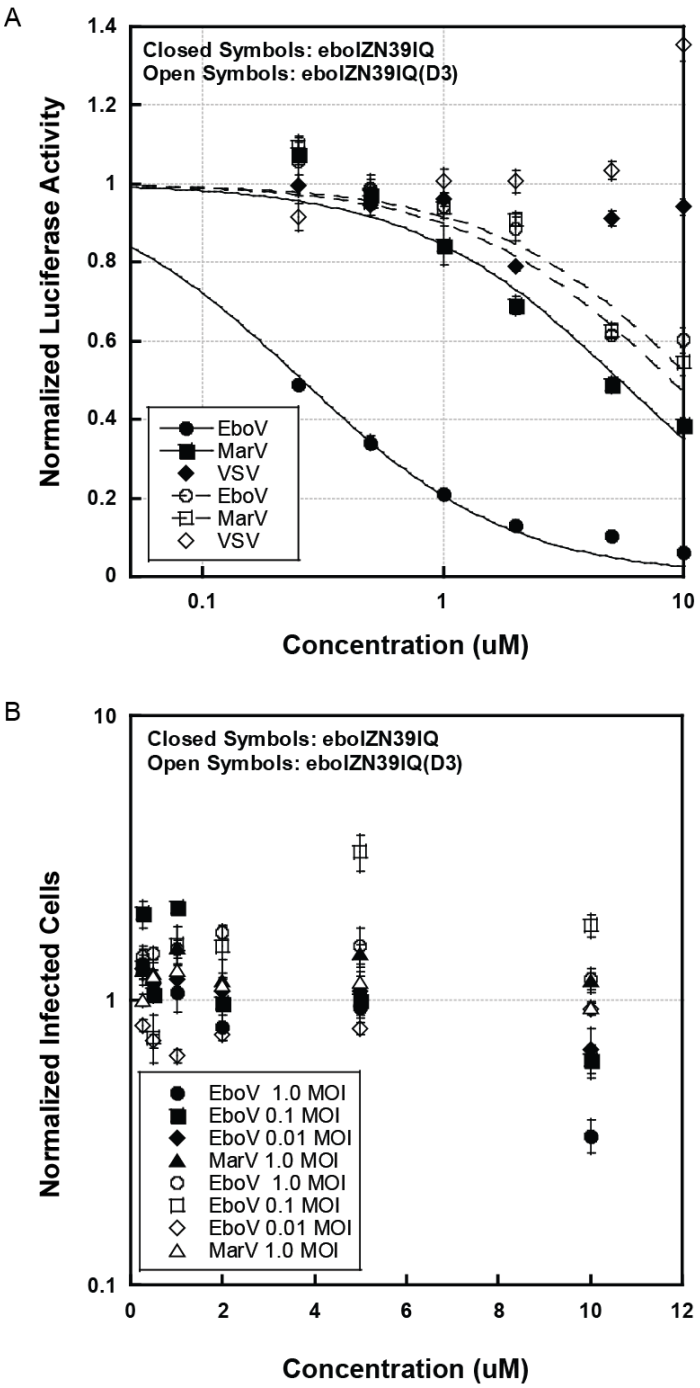


Figure 5.8. Inhibition of filovirus entry by eboIZN39IQ. A) A representative pseudovirion assay looking at the inhibitory activity of eboIZN39IQ and the negative control, eboIZN39IQ(D3) against EboV, MarV, and VSV retroviral pseudotypes. Each point represents the average of quadruplicate measurements normalized to uninhibited control. Error bars represent normalized standard errors. For this particular assay, eboIZN39IQ IC_{50}s are 260 nM against EboV and 5.4 μM against MarV. B) Data for the authentic filovirus immunofluorescence inhibition assay. Each point represents the average of quadruplicate measurements normalized to vehicle control.

Table 5.1. CD analysis of eboV N-trimer mimics

Peptide	$\theta_{222\ nm}$ (deg cm^2 $dmol^{-1}$) 25 °C	$\theta_{222\ nm}$ (deg cm^2 $dmol^{-1}$) 37 °C	$\theta_{222\ nm}$ (deg cm^2 $dmol^{-1}$) 50 °C	M_{obs}/M_{calc} 4 °C
eboIZN39IQ	-29,351	-27,910	-27,083	3.19
eboIZN39IQ(D3)	-30,369	-29,348	-28,353	3.22
eboIZN21	-25,520	-23,989	-22,868	3.45
eboIZN21(D2)	-24,849	-22,762	-21,847	3.16

CD scans were performed on the same samples at 11.4 uM eboIZN39IQ, 11.1 uM eboIZN39IQ(D3), 18.0 uM eboIZN21, and 25.3 uM eboIZN21(D2) in 50 mM Sodium phosphate, pH 5.8, 150 mM NaCl at 25 °C, 37 °C, and 50 °C. The peptides were allowed to equilibrate at each temperature for 10 min, after which no change in signal was seen over time. Sedimentation equilibrium analysis was performed on each peptide at three concentrations each (a starting concentration and two 2-fold dilutions, with typical starting concentrations between 10-30 μM) and a minimum of two speeds, but typically three speeds (18,000, 21,000, and 24,000 RPM). Each data set was globally fit to a single ideal species with a nonlinear least squares algorithm. Each sedimentation equilibrium analysis was performed 1 to 3 times and averaged for the above table.

References

1. Sanchez Aea. Fields Virology. Knipe DMH, P.M., editor. Philadelphia: Lippincott, Williams, & Wilkins; 2001. 1279-304 p.

2. Prevention CfDCa. Known cases and outbreaks of Ebola hemorrhagic fever, in chronological order 2012. Available from: http://cdc.gov/ncidod/dvrd/spb/mnpages/dispages/ebola/ebolatable.htm.

3. Bossi P, Garin D, Guihot A, Gay F, Crance JM, Debord T, et al. Bioterrorism: management of major biological agents. Cell Mol Life Sci. 2006;63(19-20):2196-212. PubMed PMID: 16964582.

4. Volchkov VE, Feldmann H, Volchkova VA, Klenk HD. Processing of the Ebola virus glycoprotein by the proprotein convertase furin. Proceedings of the National Academy of Sciences of the United States of America. 1998;95(10):5762-7. Epub 1998/05/20. PubMed PMID: 9576958; PubMed Central PMCID: PMC20453.

5. Volchkov VE, Volchkova VA, Stroher U, Becker S, Dolnik O, Cieplik M, et al. Proteolytic processing of Marburg virus glycoprotein. Virology. 2000;268(1):1-6. Epub 2000/02/23. doi: 10.1006/viro.1999.0110. PubMed PMID: 10683320.

6. Carette JE, Raaben M, Wong AC, Herbert AS, Obernosterer G, Mulherkar N, et al. Ebola virus entry requires the cholesterol transporter Niemann-Pick C1. Nature. 2011;477(7364):340-3. Epub 2011/08/26. doi: 10.1038/nature10348. PubMed PMID: 21866103; PubMed Central PMCID: PMC3175325.

7. Chandran K, Sullivan NJ, Felbor U, Whelan SP, Cunningham JM. Endosomal proteolysis of the Ebola virus glycoprotein is necessary for infection. Science. 2005;308(5728):1643-5. Epub 2005/04/16. doi: 10.1126/science.1110656. PubMed PMID: 15831716.

8. Cote M, Misasi J, Ren T, Bruchez A, Lee K, Filone CM, et al. Small molecule inhibitors reveal Niemann-Pick C1 is essential for Ebola virus infection. Nature. 2011;477(7364):344-8. Epub 2011/08/26. doi: 10.1038/nature10380. PubMed PMID: 21866101; PubMed Central PMCID: PMC3230319.

9. Misasi J, Chandran K, Yang JY, Considine B, Filone CM, Cote M, et al. Filoviruses require endosomal cysteine proteases for entry but exhibit distinct protease preferences. Journal of virology. 2012;86(6):3284-92. Epub 2012/01/13. doi: 10.1128/JVI.06346-11. PubMed PMID: 22238307; PubMed Central PMCID: PMC3302294.

10. Schornberg K, Matsuyama S, Kabsch K, Delos S, Bouton A, White J. Role of endosomal cathepsins in entry mediated by the Ebola virus glycoprotein. Journal of virology. 2006;80(8):4174-8. Epub 2006/03/31. doi: 10.1128/JVI.80.8.4174-4178.2006. PubMed PMID: 16571833; PubMed Central PMCID: PMC1440424.

11. Eckert DM, Kim PS. Mechanisms of viral membrane fusion and its inhibition. Annu Rev Biochem. 2001;70:777-810. PubMed PMID: 11395423.

12. Lee JE, Saphire EO. Ebolavirus glycoprotein structure and mechanism of entry. Future Virol. 2009;4(6):621-35. Epub 2009/01/01. doi: 10.2217/fvl.09.56. PubMed PMID: 20198110; PubMed Central PMCID: PMC2829775.

13. White JM, Schornberg KL. A new player in the puzzle of filovirus entry. Nat Rev Microbiol. 2012;10(5):317-22. Epub 2012/04/12. doi: 10.1038/nrmicro2764. PubMed PMID: 22491356; PubMed Central PMCID: PMC3540776.

14. Francis JN, Redman JS, Eckert DM, Kay MS. Design of a Modular Tetrameric Scaffold for the Synthesis of Membrane-Localized d-Peptide Inhibitors of HIV-1 Entry. Bioconjug Chem. 2012. Epub 2012/05/02. doi: 10.1021/bc300076f. PubMed PMID: 22545664; PubMed Central PMCID: PMC3429785.

15. Welch BD, Francis JN, Redman JS, Paul S, Weinstock MT, Reeves JD, et al. Design of a potent D-peptide HIV-1 entry inhibitor with a strong barrier to resistance. Journal of virology. 2010;84(21):11235-44. Epub 2010/08/20. doi: 10.1128/JVI.01339-10. PubMed PMID: 20719956; PubMed Central PMCID: PMC2953169.

16. Welch BD, VanDemark AP, Heroux A, Hill CP, Kay MS. Potent D-Peptide Inhibitors of HIV-1 Entry. Proc Natl Acad Sci U S A. 2007;104(43):16828-33. PubMed PMID: 8846219.

17. Root MJ, Steger HK. HIV-1 gp41 as a target for viral entry inhibition. Curr Pharm Des. 2004;10(15):1805-25. Epub 2004/06/08. PubMed PMID: 15180542.

18. Weinstock MT, Francis JN, Redman JS, Kay MS. Protease-resistant peptide design-empowering nature's fragile warriors against HIV. Biopolymers. 2012;98(5):431-42. doi: 10.1002/bip.22073. PubMed PMID: 23203688; PubMed Central PMCID: PMC3548907.

19. Zawadzke LE, Berg JM. A racemic protein. Journal of the American Chemical Society. 1992;114(10):4002-3. doi: 10.1021/ja00036a073.

20. Dintzis HM, Symer DE, Dintzis RZ, Zawadzke LE, Berg JM. A comparison of the immunogenicity of a pair of enantiomeric proteins. Proteins. 1993;16(3):306-8. PubMed PMID: 8346194.

21. Eckert DM, Malashkevich VN, Hong LH, Carr PA, Kim PS. Inhibiting HIV-1 entry: discovery of D-peptide inhibitors that target the gp41 coiled-coil pocket. Cell. 1999;99(1):103-15. Epub 1999/10/16. doi: S0092-8674(00)80066-5 [pii]. PubMed PMID: 10520998.

22. Schumacher TN, Mayr LM, Minor DL, Jr., Milhollen MA, Burgess MW, Kim PS. Identification of D-peptide ligands through mirror-image phage display. Science. 1996;271(5257):1854-7. PubMed PMID: 8596952.

23. Eckert DM, Kim PS. Design of potent inhibitors of HIV-1 entry from the gp41 N-peptide region. Proc Natl Acad Sci USA. 2001;98(20):11187-92. PubMed PMID: 11572974.

24. Eckert DM, Malashkevich VN, Kim PS. Crystal structure of GCN4-pIQI, a trimeric coiled coil with buried polar residues. J Mol Biol. 1998;284(4):859-65. Epub 1998/12/05. doi: 10.1006/jmbi.1998.2214. PubMed PMID: 9837709.

25. Malashkevich VN, Schneider BJ, McNally ML, Milhollen MA, Pang JX, Kim PS. Core structure of the envelope glycoprotein GP2 from Ebola virus at 1.9-A resolution. Proc Natl Acad Sci U S A. 1999;96(6):2662-7. PubMed PMID: 10077567.

26. Weissenhorn W, Carfi A, Lee KH, Skehel JJ, Wiley DC. Crystal structure of the Ebola virus membrane fusion subunit, GP2, from the envelope glycoprotein ectodomain. Mol Cell. 1998;2(5):605-16. PubMed PMID: 9844633.

27. Miller MD, Geleziunas R, Bianchi E, Lennard S, Hrin R, Zhang H, et al. A human monoclonal antibody neutralizes diverse HIV-1 isolates by binding a critical gp41 epitope. Proc Natl Acad Sci U S A. 2005;102(41):14759-64. PubMed PMID: 16203977.

28. Hackenberger CP, Schwarzer D. Chemoselective ligation and modification strategies for peptides and proteins. Angew Chem Int Ed Engl. 2008;47(52):10030-74. Epub 2008/12/17. doi: 10.1002/anie.200801313. PubMed PMID: 19072788.

29. Blanco-Canosa JB, Dawson PE. An efficient Fmoc-SPPS approach for the generation of thioester peptide precursors for use in native chemical ligation. Angew Chem Int Edit. 2008;47(36):6851-5. doi: Doi 10.1002/Anie.200705471. PubMed PMID: ISI:000258835300023.

30. Wan Q, Danishefsky SJ. Free-radical-based, specific desulfurization of cysteine: a powerful advance in the synthesis of polypeptides and glycopolypeptides. Angew Chem Int Ed Engl. 2007;46(48):9248-52. Epub 2007/11/30. doi: 10.1002/anie.200704195. PubMed PMID: 18046687.

31. Joshi SB, Dutch RE, Lamb RA. A core trimer of the paramyxovirus fusion protein: parallels to influenza virus hemagglutinin and HIV-1 gp41. Virology. 1998;248(1):20-34. Epub 1998/08/15. doi: 10.1006/viro.1998.9242. PubMed PMID: 9705252.

32. Liu IJ, Kao CL, Hsieh SC, Wey MT, Kan LS, Wang WK. Identification of a minimal peptide derived from heptad repeat (HR) 2 of spike protein of SARS-CoV and combination of HR1-derived peptides as fusion inhibitors. Antiviral research. 2009;81(1):82-7. Epub 2008/11/06. doi: 10.1016/j.antiviral.2008.10.001. PubMed PMID: 18983873.

33. Wild C, Greenwell T, Matthews T. A synthetic peptide from HIV-1 gp41 is a potent inhibitor of virus-mediated cell-cell fusion. AIDS Res Hum Retroviruses. 1993;9(11):1051-3. PubMed PMID: 8312047.

34. Watanabe S, Takada A, Watanabe T, Ito H, Kida H, Kawaoka Y. Functional importance of the coiled-coil of the Ebola virus glycoprotein. J Virol. 2000;74(21):10194-201. PubMed PMID: 11024148.

35. Miller EH, Harrison JS, Radoshitzky SR, Higgins CD, Chi X, Dong L, et al. Inhibition of Ebola virus entry by a C-peptide targeted to endosomes. J Biol Chem. 2011;286(18):15854-61. Epub 2011/04/02. doi: 10.1074/jbc.M110.207084. PubMed PMID: 21454542; PubMed Central PMCID: PMC3091195.

36. Schmidt N, Mishra A, Lai GH, Wong GC. Arginine-rich cell-penetrating peptides. FEBS Lett. 2010;584(9):1806-13. Epub 2009/11/21. doi: 10.1016/j.febslet.2009.11.046. PubMed PMID: 19925791.

37. Rydberg HA, Matson M, Amand HL, Esbjorner EK, Norden B. Effects of tryptophan content and backbone spacing on the uptake efficiency of cell-penetrating peptides. Biochemistry. 2012;51(27):5531-9. Epub 2012/06/21. doi: 10.1021/bi300454k. PubMed PMID: 22712882.

38. Higgins CD, Koellhoffer JF, Chandran K, Lai JR. C-peptide inhibitors of Ebola virus glycoprotein-mediated cell entry: effects of conjugation to cholesterol and side chain-side chain crosslinking. Bioorganic & medicinal chemistry letters. 2013;23(19):5356-60. Epub 2013/08/22. doi: 10.1016/j.bmcl.2013.07.056. PubMed PMID: 23962564; PubMed Central PMCID: PMC3822755.

39. Daniels TR, Bernabeu E, Rodriguez JA, Patel S, Kozman M, Chiappetta DA, et al. The transferrin receptor and the targeted delivery of therapeutic agents against cancer. Biochim Biophys Acta. 2012;1820(3):291-317. Epub 2011/08/20. doi: 10.1016/j.bbagen.2011.07.016. PubMed PMID: 21851850; PubMed Central PMCID: PMC3500658.

40. Lee JH, Engler JA, Collawn JF, Moore BA. Receptor mediated uptake of peptides that bind the human transferrin receptor. Eur J Biochem. 2001;268(7):2004-12. Epub 2001/03/30. PubMed PMID: 11277922.

41. Oh S, Kim BJ, Singh NP, Lai H, Sasaki T. Synthesis and anti-cancer activity of covalent conjugates of artemisinin and a transferrin-receptor targeting peptide. Cancer Lett. 2009;274(1):33-9. Epub 2008/10/08. doi: 10.1016/j.canlet.2008.08.031. PubMed PMID: 18838215.

42. Kuhn JH, Radoshitzky SR, Guth AC, Warfield KL, Li W, Vincent MJ, et al. Conserved receptor-binding domains of Lake Victoria marburgvirus and Zaire ebolavirus bind a common receptor. J Biol Chem. 2006;281(23):15951-8. Epub 2006/04/06. doi: 10.1074/jbc.M601796200. PubMed PMID: 16595665.

43. Connor RI, Chen BK, Choe S, Landau NR. Vpr is required for efficient replication of human immunodeficiency virus type-1 in mononuclear phagocytes. Virology. 1995;206(2):935-44. PubMed PMID: 7531918.

44. He J, Choe S, Walker R, Di Marzio P, Morgan DO, Landau NR. Human immunodeficiency virus type 1 viral protein R (Vpr) arrests cells in the G2 phase of the cell cycle by inhibiting p34cdc2 activity. J Virol. 1995;69(11):6705-11. PubMed PMID: 7474080.

45. Deng H, Liu R, Ellmeier W, Choe S, Unutmaz D, Burkhart M, et al. Identification of a major co-receptor for primary isolates of HIV-1. Nature. 1996;381(6584):661-6. PubMed PMID: 8649511.

46. Landau NR, Littman DR. Packaging system for rapid production of murine leukemia virus vectors with variable tropism. J Virol. 1992;66(8):5110-3. PubMed PMID: 1321291.

47. Studier FW. Protein production by auto-induction in high density shaking cultures. Protein Expr Purif. 2005;41(1):207-34. Epub 2005/05/26. PubMed PMID: 15915565.

48. Blanco-Canosa JB, Dawson PE. An efficient Fmoc-SPPS approach for the generation of thioester peptide precursors for use in native chemical ligation. Angew Chem Int Ed Engl. 2008;47(36):6851-5. Epub 2008/07/25. doi: 10.1002/anie.200705471. PubMed PMID: 18651678; PubMed Central PMCID: PMC3182823.

49. Edelhoch H. Spectroscopic determination of tryptophan and tyrosine in proteins. Biochemistry. 1967;6(7):1948-54. PubMed PMID: 6049437.

50. Cole JL. Analysis of heterogeneous interactions. Methods Enzymol. 2004;384:212-32. Epub 2004/04/15. doi: 10.1016/S0076-6879(04)84013-8. PubMed PMID: 15081689; PubMed Central PMCID: PMC2924680.

51. Laue T, Shah B, Ridgeway T, Pelletier S. Computer-aided interpretation of analytical sedimentation data for proteins. Analytical Ultrancentrifugation in Biochemistry and Polymer Science: Royal Society of Chemistry, Cambridge, UK; 1992.

52. Wohlfarth C. Partial specific volume of poly(ethylene glycol). In: Lechner MD, Arndt KF, editors. Polymer Solutions: Springer-Verlag Berlin Heidelberg; 2010.

53. Henikoff S, Henikoff JG. Amino acid substitution matrices from protein blocks. Proceedings of the National Academy of Sciences of the United States of America. 1992;89(22):10915-9. Epub 1992/11/15. PubMed PMID: 1438297; PubMed Central PMCID: PMC50453.

CHAPTER 6

SYNTHESIS AND FOLDING OF A 312-RESIDUE MIRROR-IMAGE ENZYME

Matthew T. Weinstock, Michael T. Jacobsen, Michael S. Kay

Introduction

All known living organisms utilize proteins composed of L-amino acids. Mirror-image D-peptides/proteins are promising therapeutic agents due to their resistance to degradation by natural proteases (1) and are also of great interest to synthetic biologists (mirror-image life) (2). Mirror-image proteins are not found in nature and are only accessible through chemical synthesis. Recent advances in chemical protein synthesis are making production of larger proteins more feasible, but a major anticipated challenge is the folding of synthetic proteins into their active conformations. *In vivo*, molecular chaperones, such as the extensively studied bacterial chaperone GroEL/ES, mediate folding and prevent aggregation of a large group of cellular proteins (3, 4). GroEL/ES is thought to interact with these diverse substrates via nonspecific hydrophobic interactions, but it is unknown whether it can recognize and fold mirror-image proteins. Here we examine the chiral specificity of GroEL/ES by studying the ability of this natural chaperone to fold a synthetic D-protein. This study required the total chemical synthesis of a 312-residue GroEL/ES-dependent protein, dapA, in both L- and D-chiralities, the longest synthetic proteins yet reported. We used these synthetic proteins to demonstrate that *E. coli* GroEL/ES folds both L- and D- forms of dapA with similar efficiency. This work extends the limits of chemical protein synthesis, reveals a previously unknown ambidextrous protein folding activity in GroEL/ES, and shows that natural chaperones can be used to fold D-proteins for drug development and mirror-image life applications.

The binding of substrates by GroEL is an intriguing instance of promiscuous molecular recognition. GroEL has been shown to interact transiently with approximately 250 cytosolic proteins in *E. coli* under normal growth conditions (5, 6). A subset of these

proteins exhibit an absolute requirement for GroEL and its cochaperone GroES to avoid aggregation and fold into their native state (6, 7). Interestingly, sequence analysis of the known set of GroEL/ES obligate substrates reveals no obvious consensus binding sequence (5), though structurally they are enriched in aggregation-prone folds (7). Furthermore, ~50% of soluble proteins from *E. coli* extracts will interact with GroEL when they are in a partially or fully unfolded conformation, but not in their native states (8).

Protein substrates trapped in non-native states have been shown to present hydrophobic surfaces that are otherwise buried in the core of the correctly folded protein. The thermodynamics of binding of these non-native states to GroEL imply a hydrophobic interaction (9), which is further supported by the hydrophobic character of GroEL apical domain residues implicated in substrate binding (10). Previous studies on the basis of substrate interaction with GroEL using short model peptides have concluded that the most important determinant of binding is the presentation of a cluster of hydrophobic residues by the substrate (11-13). The only evidence addressing the chiral specificity of GroEL/ES comes from a study that qualitatively demonstrated binding of a short D-peptide to GroEL (12). However, this NMR study required peptide concentrations that greatly exceed physiologic levels and did not localize the interaction to the substrate-binding region of GroEL. Until now, it has not been possible to directly test the stereo-specificity of the GroEL/ES folding reaction because of synthetic limitations in producing large D-proteins, leaving open the question of whether GroEL/ES substrate binding or folding activity is stereo-specific.

Due to our interest in mirror-image proteins as targets for drug discovery (14, 15) and mirror-image synthetic biology (e.g., the "*D. coli*" project, see below), we were intrigued by the possibility that GroEL/ES could assist in the folding of D-proteins. Thus, we decided to rigorously assess chiral specificity in the chaperone activity of GroEL/ES by synthesizing a D-version of a substrate protein and evaluating its folding by GroEL/ES. Furthermore, because most GroEL/ES substrate proteins are large (>250 residues), we reasoned that this project would provide an excellent opportunity to overcome the current limits of chemical protein synthesis.

Results and Discussion

We began by searching for a model protein that (a) requires GroEL/ES for folding under physiologic conditions and (b) has a robust activity assay that does not depend on complex chiral reagents (e.g., cofactors or other enzymes that would also have to be synthesized in mirror-image). The *E. coli* dapA protein (4-hydroxy-tetrahydrodipicolinate synthase, E.C. 4.3.3.7) meets these criteria (Fig. 6.1a,b). DapA is a 31 kDa protein that forms a homotetramer (16) and catalyzes the condensation of L-aspartate-β-semialdehyde and pyruvate to (4S)-4-hydroxy-2,3,4,5-tetrahydro-(2S)-dipicolinic acid (17-19), a key step in the biosynthesis of lysine and diaminopimelic acid (DAP), a cell-wall precursor.

DapA is highly enriched in GroEL/ES complexes under normal growth conditions (6) and is insoluble in GroEL-depleted cells (7). In *E. coli* GroEL depletion strains, cell death occurs via lysis due to a loss in cell wall integrity resulting from a lack of dapA activity, further demonstrating the dependence of dapA on GroEL/ES to adopt its native structure (20). Indeed, *in vitro*, dapA is absolutely dependent on GroEL/ES for proper

folding upon dilution from denaturant into physiologic buffer at 37ºC (6) (Fig. 6.2a).

Because D-proteins can only be accessed through chemical synthesis at present, a synthetic route to dapA was devised. Synthetic peptides are routinely made using solid-phase peptide synthesis (SPPS), but a project of this magnitude (312 residues) is well outside the capability of current SPPS technology (generally ~50 residues). In order to access larger synthetic assemblies, chemoselective ligation techniques are used to assemble peptide fragments into larger constructs (21). We employed a recently developed method to join peptide fragments via native peptide bonds formed between a peptide with a C-terminal hydrazide and a peptide with an N-terminal Cys (22). We selected this chemistry because of the convenient route to peptide hydrazides via Fmoc SPPS, the robustness of the ligation reaction, and the flexibility to carry out protein assembly in either N- to C- or C- to N-terminal directions (allowing for convergent as opposed to traditional linear C- to N-assembly).

Our retrosynthetic analysis began by locating all Cys residues (potential ligation junctions) in dapA (Fig. 6.1b), all of which are located at acceptable ligation junctions (see (22) for a discussion of unacceptable junctions). This information allowed us to break the protein into six fragments. We were initially concerned that the length of the second fragment (80 residues) and the fourth fragment (77 residues) would make synthesis/isolation of the correct product difficult.

To expand the range of potential ligation junctions, we utilized a free-radical-based desulfurization reaction that enables selective conversion of unprotected Cys into Ala (23). This technique allows one to mutate a native Ala residue to Cys during peptide synthesis (for use in ligation), and then convert the Cys back to the native Ala following

assembly. Using this methodology, we introduced two additional junction sites at A57 and A191, resulting in eight fragments overall, ranging in size from 27 to 50 residues. We refer to these peptides as dapA 1 through 8 (Fig. 6.3). Using highly optimized SPPS reaction conditions and RP-HPLC column selection, we synthesized and purified all eight peptides.

Our initial strategy for the assembly of these eight fragments required 12 steps and their associated purifications (seven ligations, two desulfurizations, three Acm removals; Fig. 6.4). Acm is an orthogonal Cys protecting group that prevents cyclization/polymerization of peptides containing both an activated C-terminal hydrazide and an N-terminal Cys, and can also be used to prevent Cys desulfurization. We found that while we were able to assemble the C-terminal portion (dapA 5-8), we were unable to assemble the N-terminal portion (dapA 1-4) following this scheme. A significant complication was the His thioester on dapA 2 (H56), which was highly susceptible to hydrolysis, leading to low reaction yields during the dapA 2 to 3 ligation step (Fig. 6.5). This difficulty, coupled with the large number of manipulations (and concomitant sample losses), resulted in a failure to assemble dapA 1-4 in usable yield.

In order to overcome these synthetic difficulties, we employed two strategies to reduce the complexity of our assembly scheme. First, we observed that we could simplify the original assembly strategy if we eliminated the desulfurization step necessary to convert the Cys to native Ala at the dapA 2-3 junction. Towards this end, we determined locations in our protein that would likely tolerate permanent mutation to Cys. We began by performing BLAST analysis of the *E. coli* dapA to identify the most similar homologs, aligning the top 1000 hits (>69% conservation, >49% identity) to

determine positions where Cys residues naturally occur. Fortuitously, 12% of the aligned sequences contained Cys at position 57, site of the dapA 2-3 junction (Fig. 6.6a). Next, we analyzed the dapA crystal structure to determine the likelihood of the A57C mutation to disrupt protein structure/function. The side-chain of residue 57 is surface exposed and is not in close proximity to the active site or any native Cys residues (>12 Å to the nearest Cys) (Fig. 6.6b). These results suggested that introduction of the A57C mutation would likely be well tolerated. Indeed, mutagenesis of dapA (dapA A57C) confirmed this prediction, as the mutation affected neither recombinant protein activity (Fig. 6.2b) nor its dependence on GroEL/ES for folding under physiological conditions (Fig. 6.2a).

Our second approach to reducing the number of assembly steps was to synthesize longer initial peptide fragments via SPPS, which was greatly aided by the use of pseudoproline dipeptides (known to disrupt aggregation during SPPS) (24). With optimized conditions, we successfully synthesized and purified dapA 1-2, dapA 5-6, and dapA 7-8 as single peptides, but we were unable to synthesize and purify dapA 3-4 in sufficient yield and purity (Fig. 6.7).

By combining rational mutagenesis (to eliminate desulfurization/Acm removal and associated purifications) with longer starting peptide fragments (some of which were dependent on pseudoproline dipeptides for sufficient purity/yield), we assembled L-dapA A57C from five fragments using five assembly steps (Fig. 6.8). However, due to the lack of commercially available D-pseudoproline dipeptides required for the D- syntheses of dapA 1-2 and dapA 5-6, we ultimately selected a final assembly strategy that incorporated both the A57C mutation and the dapA 7-8 unified fragment, which did not depend on pseduoprolines. This final strategy yielded a seven-fragment assembly

scheme (Figs. 6.3 and 6.9), ultimately removing four steps (and associated purifications) from the initial scheme.

Following this final simplified strategy, we successfully assembled the 312-residue synthetic dapA A57C (hereafter referred to as "dapA") in both L- and D-chiralities (Figs. 6.10, 6.11, and 6.12), the longest synthetic proteins (in both chiralities) reported to date. The proteins were synthesized at milligram scale (1.66 mg of D-dapA, total yield of 1.14%), which is a >200-fold improvement over the longest previously reported nonrepetitive synthetic protein (the 286-residue F-ATPase γ subunit, of which ~30 μg was synthesized, at <0.005% yield (25)). The synthetic L- and D-proteins behaved identically to recombinant dapA on a C4 RP-HPLC column (Fig. 6.10a) and possess the correct molecular weights (Fig. 6.10b, c)

Because our synthetic proteins are denatured during the assembly process, we developed a chemical folding protocol to fold dapA in the absence of GroEL/ES (Fig. 6.13a), providing a means to evaluate enzymatic activity of our synthetic constructs. Our method employs 0.5 M L-arginine (a protein refolding additive (26)), was validated using recombinant dapA, and works equally well with D-arginine (Fig. 6.13b). After arginine-assisted folding of synthetic L- and D-dapA, we used size exclusion chromatography (SEC) to isolate highly active tetrameric protein, removing unfolded material (Fig. 6.14). Following SEC, both the L- and D-dapA synthetic constructs have the expected CD spectra (Fig. 6.15a; note D-dapA's mirror-image spectrum was inverted to aid comparison with L-dapA) and proved to be enzymatically active (Fig. 6.15b), demonstrating that our synthetic material is correctly folded and functional.

With folded and validated synthetic L- and D-dapA in hand, we were poised to perform the definitive experiment comparing the folding of our synthetic L- and D-dapA by GroEL/ES. The proteins were denatured in 6 M GuHCl and diluted into GroEL/ES-containing buffer to initiate refolding. Both synthetic L- and D-dapA recovered significant activity after GroEL/ES-mediated refolding (Fig. 6.16). Interestingly, the refolding of D-dapA was slightly less efficient than synthetic L-dapA, which may reflect a subtle difference(s) in the interaction of the GroEL apical domain with D-dapA.

Importantly, the results presented here demonstrate that GroEL/ES is able to fold a D-protein and therefore does not manifest strict stereo-specificity in folding its substrates. This result provides strong support for a substrate binding mechanism via nonspecific hydrophobic interactions followed by sequestration in a passive Anfinsen cage (3). Our study provides proof-of-concept for the use of natural (L-) GroEL/ES to fold D-proteins for mirror-image drug discovery and synthetic biology applications. It is now possible to envision the use of a GroEL/ES-containing bioreactor for folding of mirror-image proteins (27).

We note that synthetic L- and D-dapA possess ~20% lower specific activity compared to recombinant despite having the correct molecular weight. We speculate that this difference is due to subtle defects in a fraction of the material (e.g., epimerization (28, 29)). Determining the nature of and polishing these defects will be crucial to expanding application of this work to even larger synthetic proteins.

In order to determine if the ability of GroEL/ES to fold D-proteins is universal, we are pursuing the total chemical synthesis of D-GroEL (548 residues) and D-GroES (97 residues), which will then be screened in refolding assays against a suite of well-

characterized recombinant L-substrates. The D-GroEL/ES project will also further advance synthesis methodologies for the assembly of large proteins, giving complete control over the proteome with atomic precision.

The ability to chemically synthesize proteins of interest not only serves to advance mirror-image drug discovery efforts by making larger targets available, but also provides alluring possibilities for mirror-image synthetic biology (2). It also complements other efforts to synthesize large biomolecules (e.g., synthetic genomes (30)). An intriguing prospect is the assembly of a mirror-image *in vitro* translation apparatus (D-ribosome) and eventually a mirror-image organism, an effort that we have dubbed the "*D. coli*" project (14). Such an organism would not only provide a facile route to mirror-image biomolecules, but would also have the added benefit of a "chiral firewall" that would prevent DNA exchange (horizontal gene transfer) and infection by other organisms (e.g., phage). A mirror-image organism would likely be unable to survive outside the lab due to the lack of necessary mirror-image nutrients in nature (a hypothesis we are currently investigating by exposing *E. coli* to a simulated mirror-image world). This organism could also permit the study of highly pathogenic agents in mirror-image, which have the same physiochemical properties as their enantiomeric counterparts without the toxicity of the naturally occurring forms. In the short term, synthetic mirror-image proteins will facilitate structural/biochemical studies of highly toxic proteins and mirror-image drug development.

Full Methods

Recombinant dapA proteins

The coding region of dapA was PCR amplified from *E. coli* BL21 (DE3) (Novagen) genomic DNA and inserted into the pET14b vector (Novagen) between NdeI and BamHI restriction sites, yielding a construct with an N-terminal hexahistidine tag and thrombin cleavage site (pET14b_dapA). pET14b_dapA_A57C (a construct with the A57C mutation) was generated by mutating *dapA* codon 57 from GCT to TGC using the Q5® Site-Directed Mutagenesis Kit (NEB) with pET14b_dapA as the template, and was confirmed by sequencing of the entire gene. Proteins were expressed in BL21 (DE3) cells (Novagen). Cultures were grown in autoinduction media (31) in shake flasks at 37ºC to an OD_{600} of 0.7-1 and then allowed to grow for an additional 14-18 hrs at 19ºC. Cell pellets were suspended in buffer A (50 mM Tris pH 7.3, 300 mM NaCl, 5 mM imidazole, 5% glycerol) and disrupted by sonication. Samples were clarified by centrifugation, and protein was isolated from the supernatant by applying to His-Select Ni-affinity resin (Sigma), thoroughly washing with buffer A, and eluting in buffer A containing 250 mM imidazole. Purified fractions were pooled, dialyzed against buffer A without imidazole, spin concentrated to ~8 mg/mL, aliquoted, flash frozen, and stored at -80ºC. A portion of the material was further purified via reverse-phase HPLC and lyophilized for use in refolding studies.

GroEL and GroES

GroEL and GroES were expressed in *E. coli* DH5α from the pBRE-groESL+ plasmid (F. Keppel and C. Georgopoulos, unpublished), which contains the *E. coli groE* operon. Cultures were grown at 37ºC in shake flasks to an OD_{600} of 0.6. An equal

volume of 55ºC media was added and the cultures were shifted to 43ºC for 3 hrs to induce expression of *groE*. Cell pellets were resuspended in buffer (50 mM Tris pH 8, 1 mM EDTA, 1 mM DTT) containing 0.17 mg/mL PMSF and lysed by incubation with lysozyme, followed by sonication and centrifugation. GroEL was purified as described in (32). GroES was essentially purified as described in (28) with the following modifications. A 5 mL HiTrap Q HP column (GE Healthcare) was used instead of a DE-52 column. Eluted fractions containing GroES were dialyzed at 4ºC into 25 mM Tris (pH 8 at 4ºC), 0.5 mM EDTA, 1 mM DTT. 1 M NaOAc pH 4.6 was added to 75 mM and the sample was stirred on ice for 15 min, filtered through a 0.2 μm pore size membrane, and then purified on a 1 mL HiTrap SP XL column (GE Healthcare) with a NaCl gradient (0-500 mM over 15 column volumes) in 50 mM NaOAc, 0.5 mM EDTA, 2mM DTT, pH 4.6. Fractions containing GroES as judged by SDS-PAGE and LC/MS analysis were pooled, dialyzed against 25 mM Tris pH 7.5 (pH 8 at 4ºC), 0.5 mM EDTA, 1 mM DTT, and concentrated in an Amicon Ultra-15 centrifugal concentrator. Concentrations were determined by absorbance at 280 mm (GroES) or Bradford method (GroEL). Glycerol was added to 10%, and aliquots were flash frozen and stored at -80ºC.

Peptide synthesis

Peptides were synthesized via Fmoc-SPPS on a Protein Technologies, Inc. Prelude peptide synthesizer. Pseudoproline dipeptides were obtained from aapptec. Multiple batches of each peptide were synthesized at 32 μmol scale. C-terminal peptides (dapA 8 and dapA 7-8) were synthesized on TentaGel R RAM resin (0.19 mmol/g, RAPP Polymere). All other peptides were synthesized on 2-hydrazine chlorotrityl resin (ChemPep) loaded manually to a density of 0.05 - 0.1 mmol/g. 2-hydrazine chlorotrityl

resin was reproducibly loaded to desired density by dissolving 0.075 mmol amino acid and 0.075 mmol Oxyma Pure (Novabiochem) in 1 mL 1:1 DMF:DCM, and activating with 0.08 DIC for 10 min before addition to 180 mg of resin. After a 2 hr coupling, the resin is washed with DMF, and unreacted groups on the resin are capped with 1:1 acetic anhydride:0.6 M NMM in DMF for 15 min. Automated synthesis was performed with the following parameters: Deprotection: 20% piperidine/DMF for 3 x 3 min; Coupling: [1.3/1.3/1] of [200 mM Fmoc-protected amino acid in NMP/198 mM HATU in DMF/600 mM NMM in DMF] for 25 min; Washing: 6 x 30 sec with DMF; Cleavage: [92.5/2.5/2.5/2.5] of [TFA/water/TIS/EDT] (15 mg/ml NH_4I is included in the cocktail for methionine-containing peptides) for 2.5 hrs. The cocktail is isolated from the resin, concentrated, and precipitated/washed with cold diethyl ether.

Peptide ligation

For a typical ligation, the N-terminal peptide fragment was dissolved (0.8-3 mM) in activation buffer (6 M GuHCl, 100 mM sodium phosphate, pH 3) and chilled to -20ºC. Hydrazide activation was accomplished by adding sodium nitrite to a final concentration of 5-20 mM (lower concentrations were found to diminish hydrolysis of dapA 2 thioester) from a 200 mM stock in water (pH adjusted to 4), mixing, and incubating at -20ºC for 20-30 min. The C-terminal peptide fragment was dissolved (~2x molar excess of N-terminal fragment, except in the case of dapA 5-8 where an excess of dapA 1-4 was used) in ligation buffer (6 M GuHCl, 200 mM 4-mercaptophenylacetic acid, 200 mM sodium phosphate, pH 7). In ligations involving dapA 2, 3, and 6, the GuHCl concentration was increased to 8 M to enhance peptide solubility. Following activation of the N-terminal fragment, the solutions containing N- and C-terminal fragments were

combined, pH adjusted to 7 with NaOH with rapid mixing, and allowed to react at 25ºC for 5-20 hrs. Following ligation, TCEP was added, without pH adjustment, to ~130 mM and incubated for 10 min. The volume was increased to 2.5 mL with 6 M GuHCl in 5% AcOH, then brought to 5 mL with 5% AcOH. The sample was sonicated, clarified via centrifugation, and purified by HPLC.

Desulfurization

Desulfurization was based on (23) with the modification of replacing *t*BuSH with glutathione (33). Desulfurization buffer (6 M GuHCl, 100 mM phosphate, pH 6.5) was sparged with He and used to prepare a 550 mM stock of TCEP (pH re-adjusted with NaOH to 6.5 after the addition of TCEP) and a 53 mM stock of reduced glutathione. A 240 mM stock of VA-044 was prepared in sparged water. Peptide was dissolved in 230 µL of desulfurization buffer at a concentration of ~3 mM. Reagents were added to the peptide solution in the following order: 204 µL glutathione stock, 204 µL VA-044 stock, 720 µL TCEP stock. The reaction was layered with argon and incubated on a tube rotisserie overnight at 37ºC. The reaction was diluted to 2.5 mL with 6 M GuHCl in 5% AcOH, mixed, then diluted to 5 mL with 5% AcOH. The peptide was purified via HPLC.

Acm deprotection

Acm deprotection was performed as described in (34) by adding AgOAc to HPLC fractions containing Acm-protected peptide to a final concentration of 20 mM, covering with argon, and incubating overnight with gentle agitation at 25ºC. DTT was added to 24 mM to quench the reaction, which resulted in an immediate metal-DTT precipitate. Samples were clarified via centrifugation, diluted with 0.1% TFA in water (to reduce acetonitrile concentration), and purified via HPLC.

HPLC and LC/MS

Analytical reverse-phase HPLC was performed on Phenomenex Jupiter 4 μm Proteo C12 90 Å (150 x 4.6 mm) and Phenomenex Jupiter 5 μm C4 300 Å (150 x 4.6 mm) columns at 1 mL/min with a water/acetonitrile gradient in 0.1% TFA. Preparative reverse-phase HPLC was performed on either Phenomenex Jupiter 4 μm Proteo C12 90 Å (250 x 21.20 mm) or Phenomenex Jupiter 10 μm C4 300 Å (250 x 21.20 mm) at 10 mL/min with a water/acetonitrile gradient in 0.1% TFA (for crude peptides), or on a Phenomenex Jupiter 10 μm C4 300 Å (250 x 10 mm) at 5 mL/min with a water/acetonitrile gradient in 0.1% TFA (for ligation products). Mobile phase was removed via lyophilization. LC/MS was performed with a Phenomenex Aeris WIDEPORE 3.6 μm C4 (50 x 2.1 mm) column at 0.4 mL/min with a water/acetonitrile gradient in 0.1% formic acid on an AB Sciex API 3000 LC/MS/MS system.

Arginine-assisted refolding

dapA constructs were dissolved or buffer exchanged into denaturation buffer (6 M GuHCl, 20 mM MOPS pH 7.4, 100 mM KCl, 10 mM MgCl2, 10 mM DTT) with 0.5 M arginine, equilibrated to 13ºC, and dialyzed (Slide-A-Lyzer mini dialysis cassettes 3500 MWCO) against 100x volume of refolding buffer (20 mM MOPS pH 7.4, 100 mM KCl, 10 mM $MgCl_2$, 10 mM sodium pyruvate, 1 mM DTT) with 0.5 M arginine for 2.5 hrs at 13ºC. The sample was then further dialyzed against 100x volume 100 mM ammonium bicarbonate for 1 hr. The dialyzed sample was used directly in assays or further purified by size exclusion chromatography (Superdex 200 10/30, GE Healthcare) in 100 mM ammonium bicarbonate (pH 8) running buffer with a flow rate of 0.75 mL/min.

dapA refolding assay

The dapA refolding assay was adapted from (6). 25 μM stocks of dapA proteins were prepared from lyophilized powder or buffer exchanged into denaturation buffer (20 mM MOPS pH 7.4, 100 mM KCl, 10 mM $MgCl_2$, 10 mM DTT, 6 M GuHCl) and allowed to denature for 1 hr at 25ºC. Refolding was initiated by diluting 100x into 37ºC refolding buffer (20 mM MOPS pH 7.4, 100 mM KCl, 10 mM $MgCl_2$, 10 mM sodium pyruvate, 5 mM ATP) with or without 7 μM GroEL monomer +/- 7 μM GroES monomer. At specific time points, 10 μL aliquots of the refolding reaction were added to 240 μL of assay buffer (200 mM Imidazole pH 7.4, 35 mM Na Pyruvate, 4 mM DL-aspartate-β-semialdehyde, 0.5 mg/mL o-aminobenzaldehyde, 12.5 mM CDTA), which simultaneously quenches chaperone-mediated refolding and initiates assay of enzyme activity. The assay is quenched after 15 min of agitation on a microplate shaker (800 RPM) by the addition of 50 μL 2 M HCl, developed by continuing the agitation for 1 hr., followed by measuring the A_{562}.

DL-Aspartate-β-semialdehyde

DL-aspartate-β-semialdehyde for use in refolding assays was prepared essentially as described by Black and Wright (35) from DL-allylglycine except that the purification on Dowex resin was omitted and the compound was aliquoted and stored in 1 M HCl at -80ºC. Stocks were neutralized with NaOH immediately before preparation of assay buffer.

Circular dichroism

All CD spectra were recorded on an AVIV Model 410 spectrophotometer (AVIV, Lakewood, NJ) in 100 mM ammonium bicarbonate buffer in a 1 mm QS quartz cuvette (Starna) at 25ºC. Wavelength scans were performed at 1 nm resolution with 1 second averaging time. Data from triplicate scans were averaged, blank subtracted, and normalized to mean residue ellipticity by the following equation: $[\theta] = 100*\theta/C*l*(n-1)$ where C is concentration of protein in mM, l is pathlength in cm, and n is the number of residues in the protein.

Sequence analysis

BLAST analysis was performed using the BLASTp algorithm with default parameters with the dapA protein sequence from *E. coli* BL21(DE3) (GenBank: ACT44191.1). The sequences were aligned with the COBALT Constraint-based Multiple Protein Alignment Tool (36) and analyzed with Jalview (37).

Structural analysis

The crystal structure of dapA (PDB ID: 1DHP) was analyzed using the Pymol Molecular Graphics System, Schrödinger, LLC.

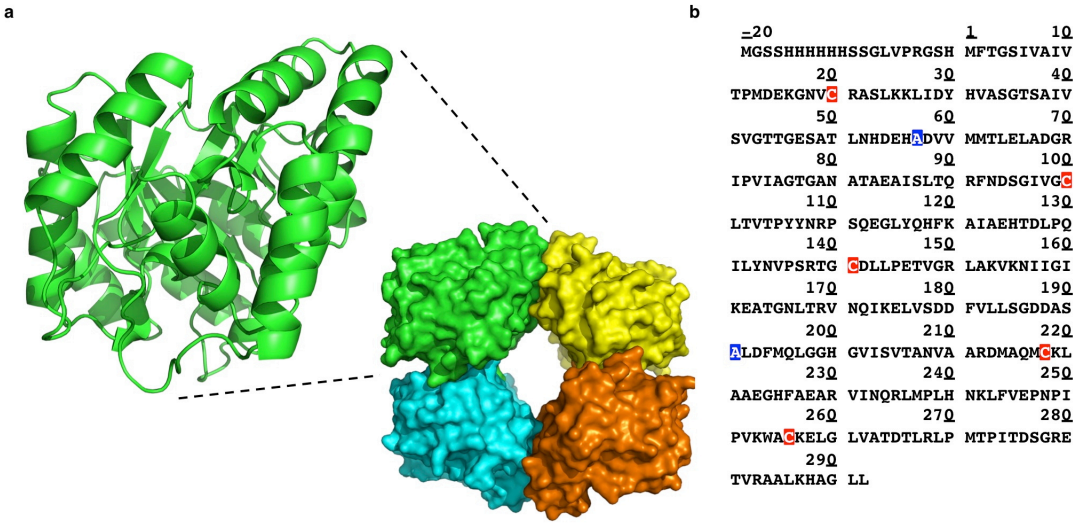


Figure 6.1. Sequence and structure of dapA. a, Sequence of our 312-residue dapA construct. Native dapA (292-residues) was prepended with an N-terminal hexahistidine tag with a thrombin cleavage site (residues -20 to -1). All native Cys residues (red) and two Ala residues used as ligation junctions (blue) are highlighted. **b**, Crystal structure of dapA (PDB ID: 1DHP), with one dapA monomer shown in green as a ribbon diagram.

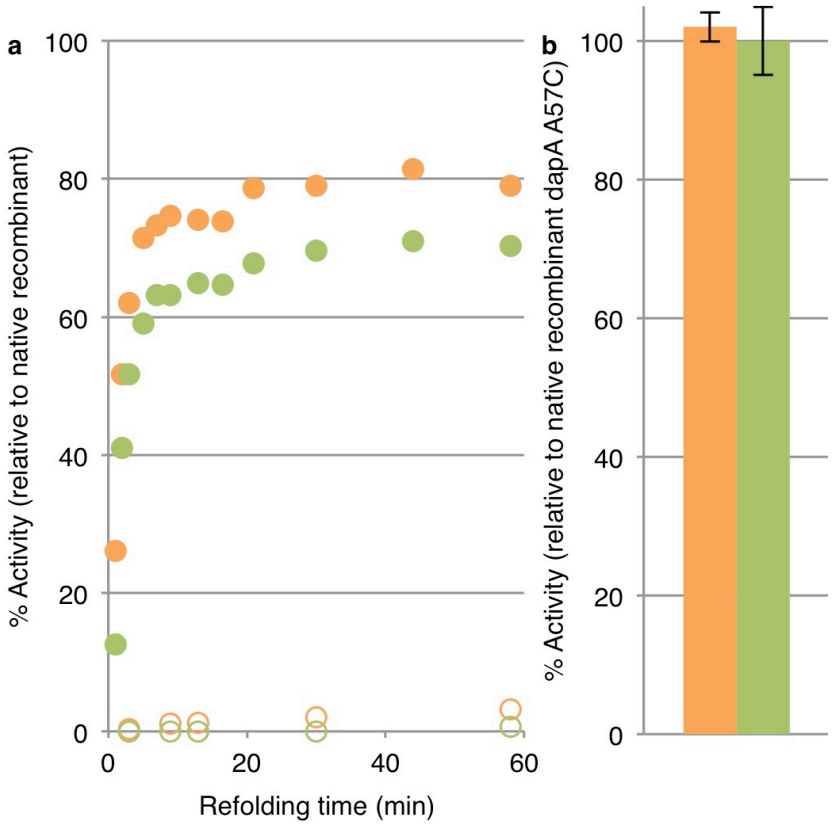


Figure 6.2. Comparison of dapA wt and dapA A57C. a, refolding of dapA wt and dapA A57C in the presence or absence of GroEL/ES. **b,** Enzymatic activity of native dapA wt and dapA A57C. Error bars indicate sd.

Peptide	residues	Sequence
dapA 1	-20 to 19	MGSSHHHHHHSSGLVPRGSHMFTGSIVAIVTPMDEKGNV-NHNH2
dapA 2	20 to 56	CRASLKKLIDYHVASGTSAIVSVGTTGESATLNHDEH-NHNH2
dapA 3	57 to 99	CDVVMMTLELADGRIPVIAGTGANATAEAISLTQRFNDSGIVG-NHNH2
dapA 4	100 to 140	CLTVTPYYNRPSQEGLYQHFKAIAEHTDLPQILYNVPSRTG-NHNH2
dapA 5	141 to 190	CDLLPETVGRLAKVKNIIGIKEATGNLTRVNQIKELVSDDFVLLSGDDAS-NHNH2
dapA 6	191 to 217	CLDFMQLGGHGVISVTANVAARDMAQM-NHNH2
dapA 7	218 to 255	CKLAAEGHFAEARVINQRLMPLHNKLFVEPNPIPVKWA-NHNH2
dapA 8	256 to 292	CKELGLVATDTLRLPMTPITDSGRETVRAALKHAGLL-amide
dapA 7-8	218 to 292	CKLAAEGHFAEARVINQRLMPLHNKLFVEPNPIPVKWACKELGLVATDTLRLPMTPITDSGRETVRAALKHAGLL-amide

Figure 6.3. Peptide fragments used in initial and final assemblies. Peptides, corresponding residue numbers (numbered according to Fig. 6.1b), and sequences of fragments used in initial and final assembly strategies. Native Cys (red), A57C mutation (purple), Cys residues introduced at native alanine sites for ligation that were later converted back to alanine through desulfurization (blue), and locations where pseudoproline dipeptides were incorporated (green) are highlighted.

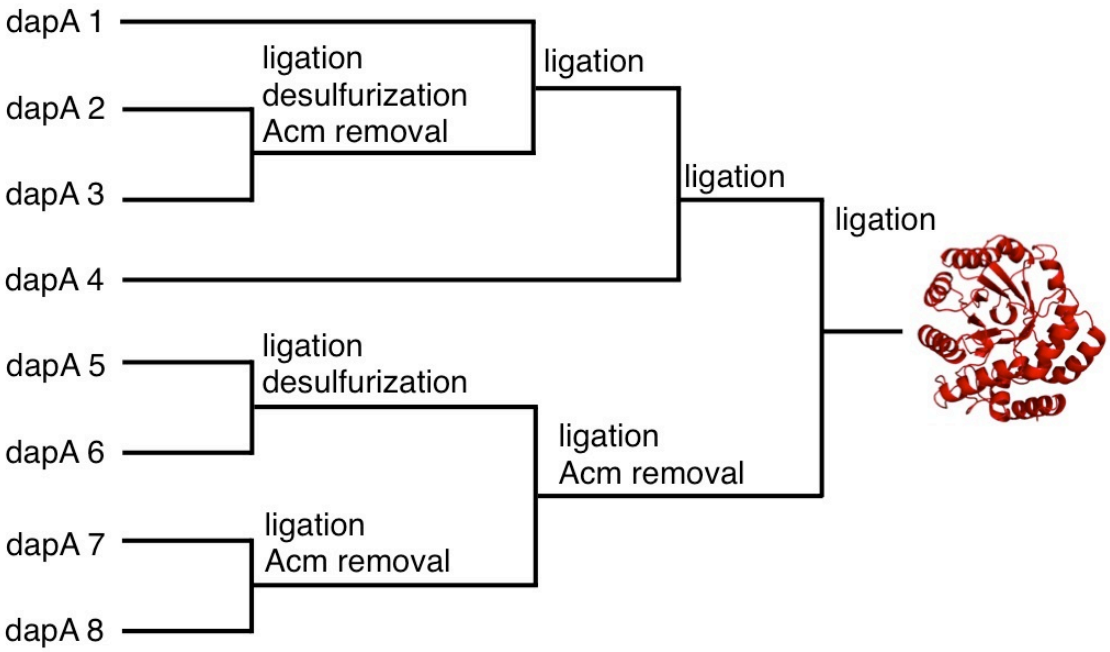


Figure 6.4. Initial dapA assembly strategy. The initial assembly scheme employed in the synthesis of L-dapA A57C (PDB structure 1DHP).

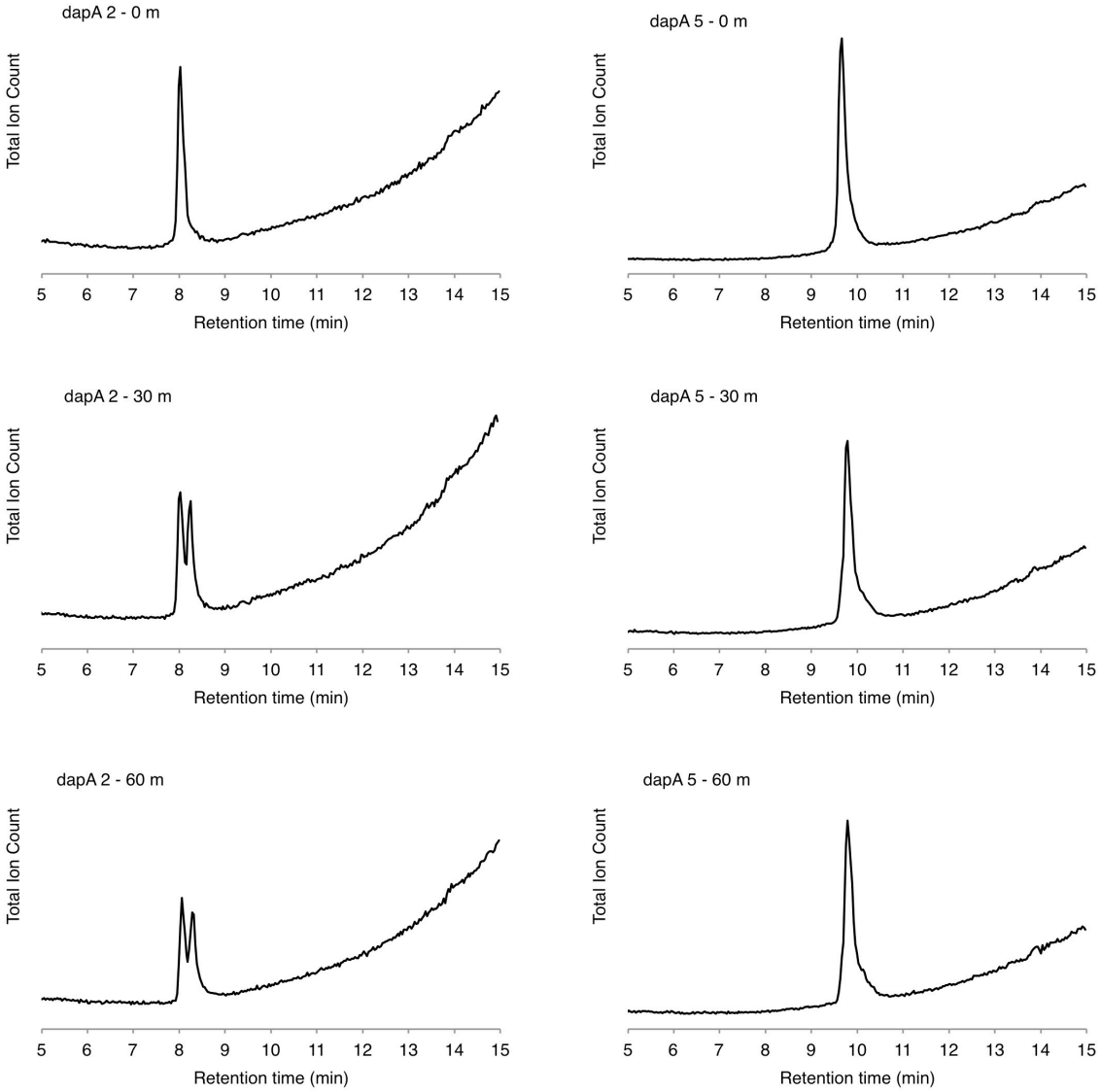


Figure 6.5. dapA 2 thioester hydrolysis. The left column shows three points of a time course of monitoring dapA 2-MPAA (an active thioester) via LC/MS. The appearance of a second peak in the 30-min time point is due to hydrolysis from dapA 2-MPAA to dapA 2-COOH. For comparison, the stability of dapA 5-MPAA is presented in the right column. No significant hydrolysis is noted after O/N incubation for dapA 5-MPAA whereas dapA 2-MPAA is fully hydrolyzed (data not shown).

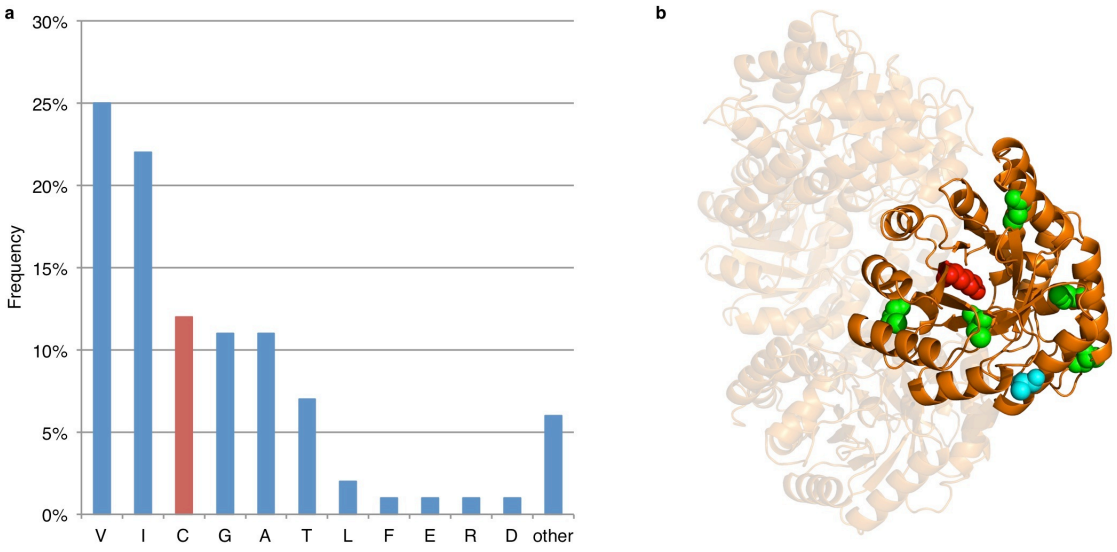


Figure 6.6. Sequence and structural analysis of the dapA A57C mutation. a, Natural sequence diversity at position 57 from the BLAST analysis. **b**, Structure of dapA tetramer (PDB: 1DHP) showing (on one subunit) the surface-exposed alanine at position 57 (cyan). Natural cysteine residues (green) and the catalytic K161 residue in the active site (red) are also shown.

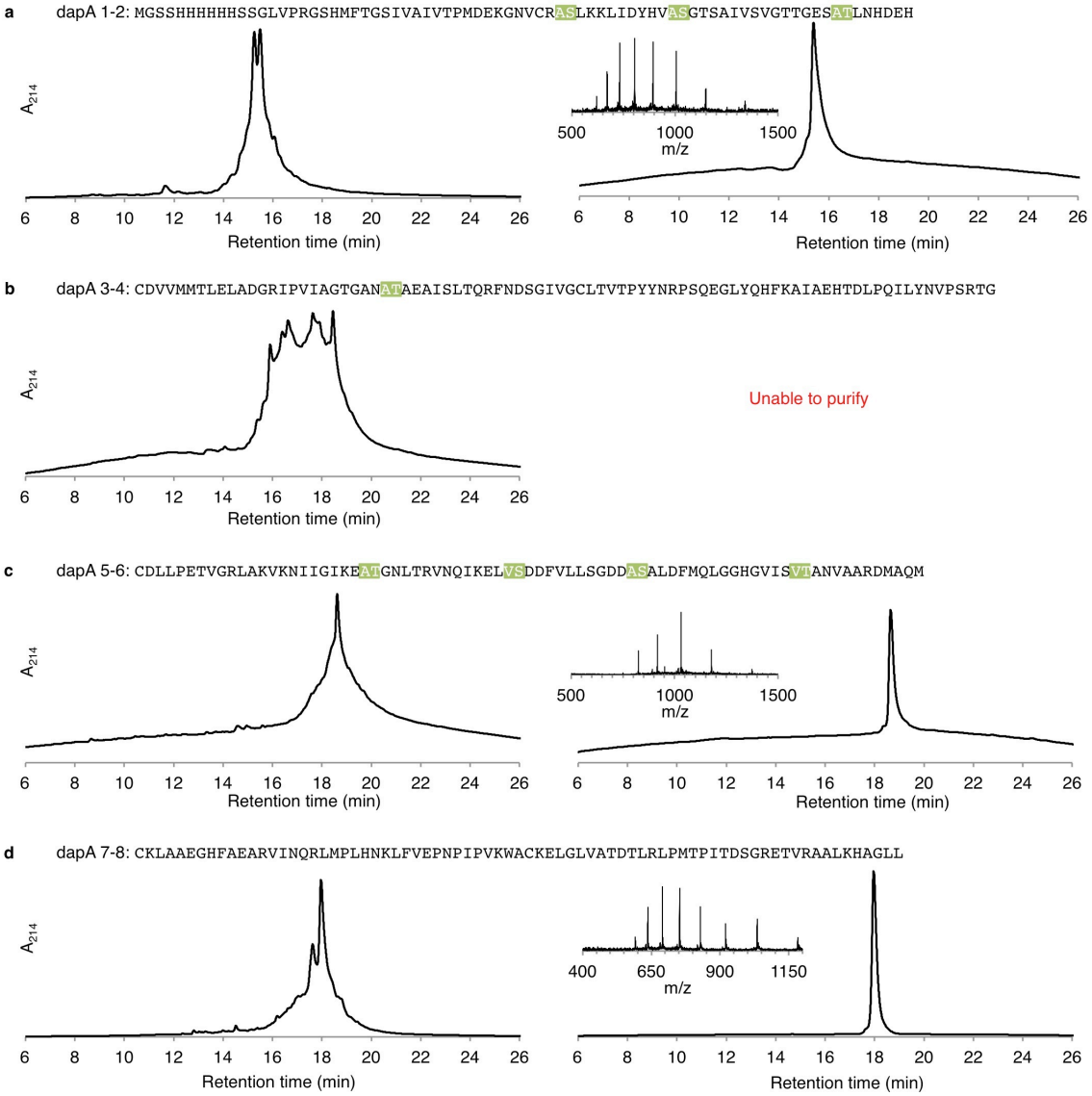


Figure 6.7. Longer initial peptide fragments via SPPS. Analytical C4 RP-HPLC traces of crude (left) and purified (right) longer dapA fragments synthesized via SPPS. **a,** dapA 1-2. **b,** dapA 3-4. **c,** dapA 5-6. **d,** dapA 7-8. Mass spectra inset on purified chromatograms. Locations of pseudoproline dipeptides highlighted in green.

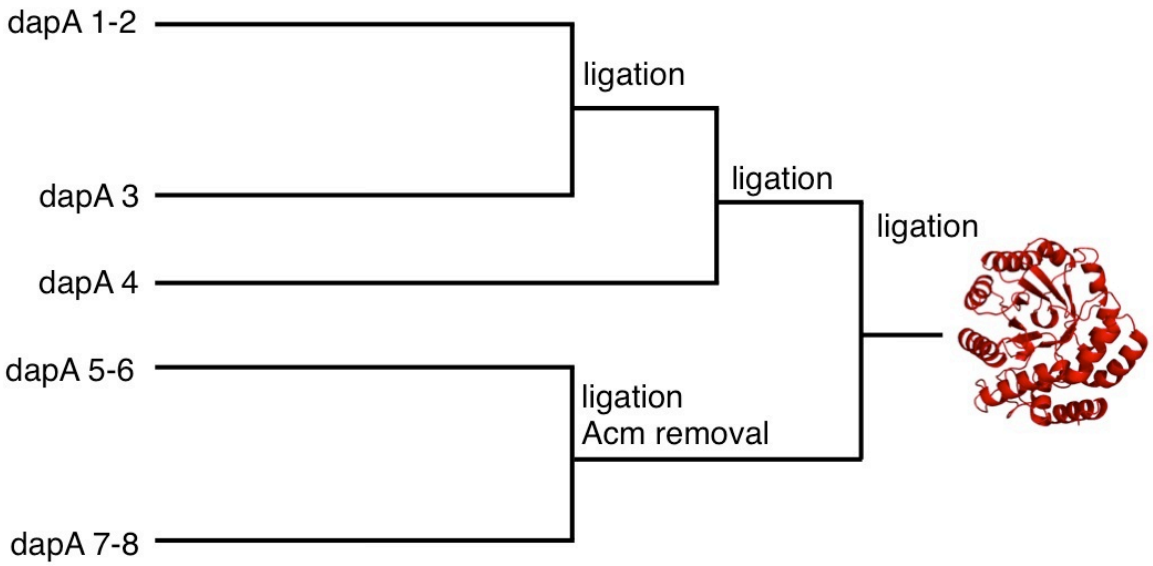


Figure 6.8. Streamlined dapA assembly strategy. This simplified assembly of synthetic L-dapA depends on incorporation of the A57C mutation and longer starting peptides.

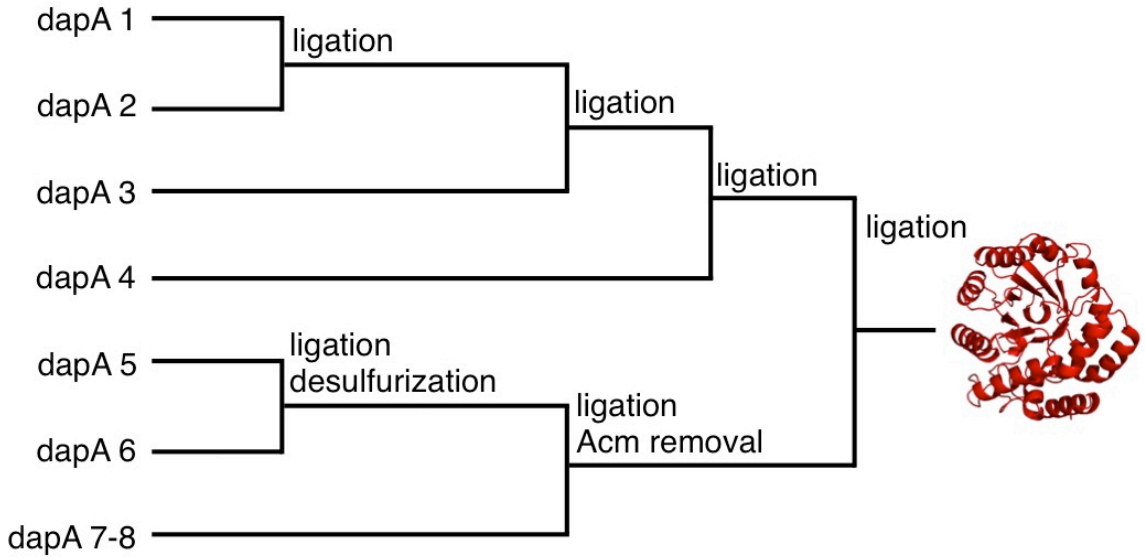


Figure 6.9. Synthetic dapA assembly. Final assembly scheme employed in the synthesis of L- and D-dapA A57C (PDB structure 1DHP).

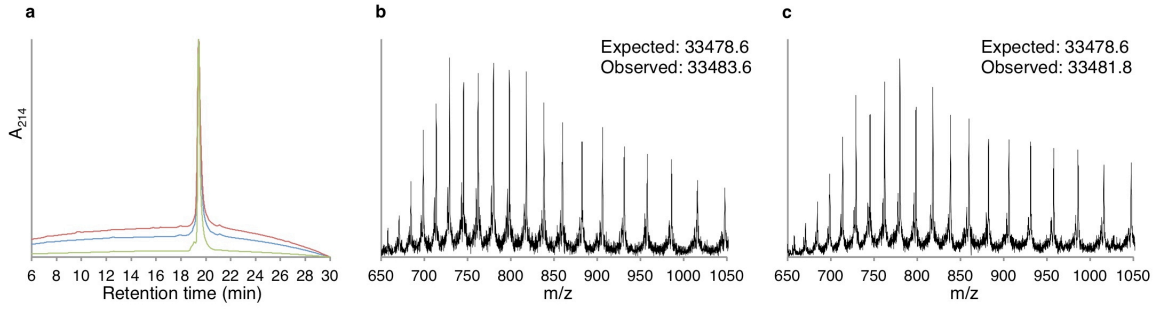


Figure 6.10. Characterization of synthetic dapA. a, analytical C4 reverse-phase HPLC of purified recombinant and synthetic dapA. **b-c**, mass spec data for synthetic L-/D-dapA.

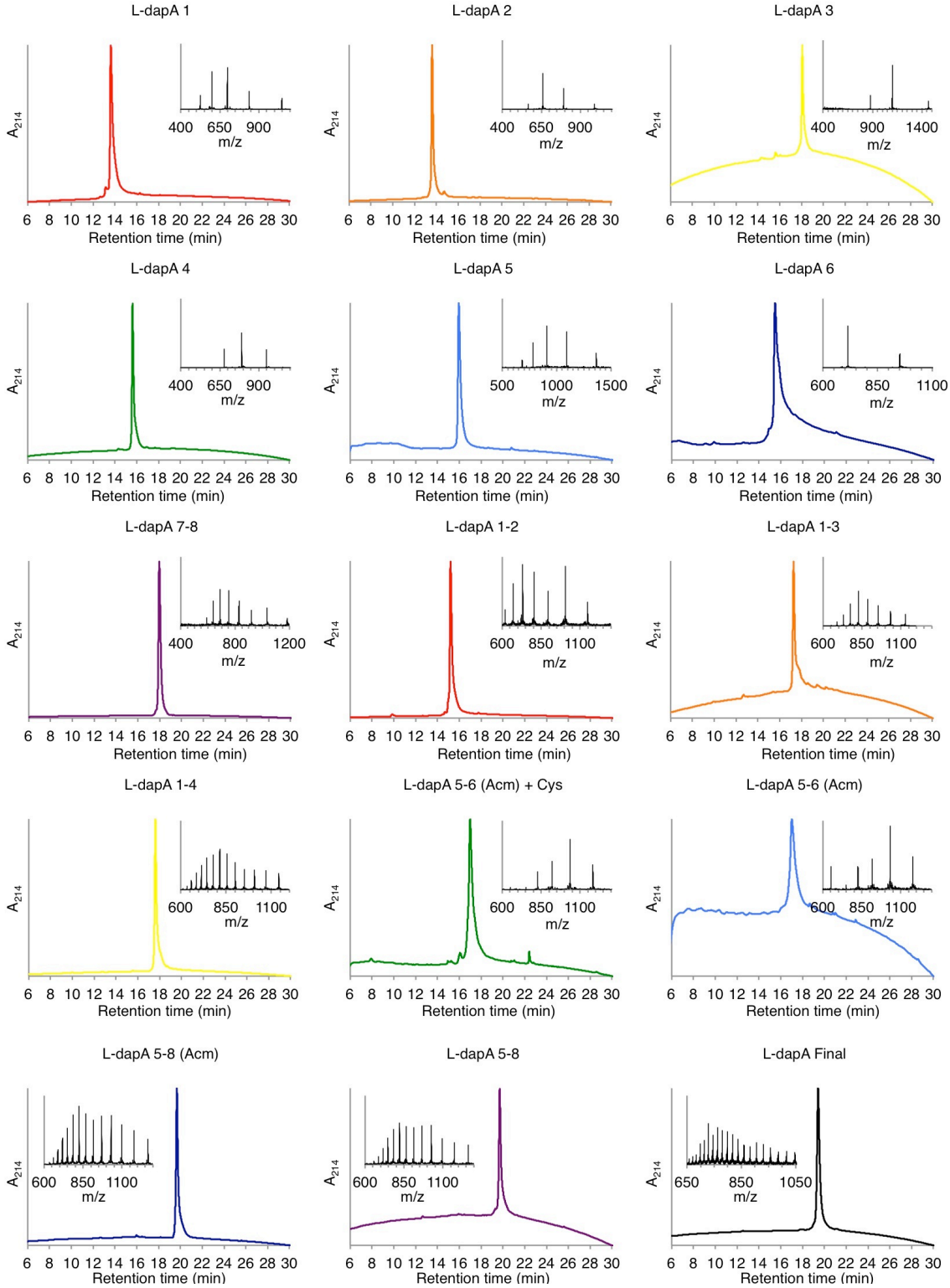


Figure 6.11. Characterization of assembly intermediates of L-dapA. Analytical RP-HPLC traces and mass spectra of individual peptides and ligation intermediates constructed for the synthesis of L-dapA. All traces were obtained on a C4 column, except dapA 5, dapA 5-6 (Acm), and desulfurized dapA 5-6 (Acm) (C12 column).

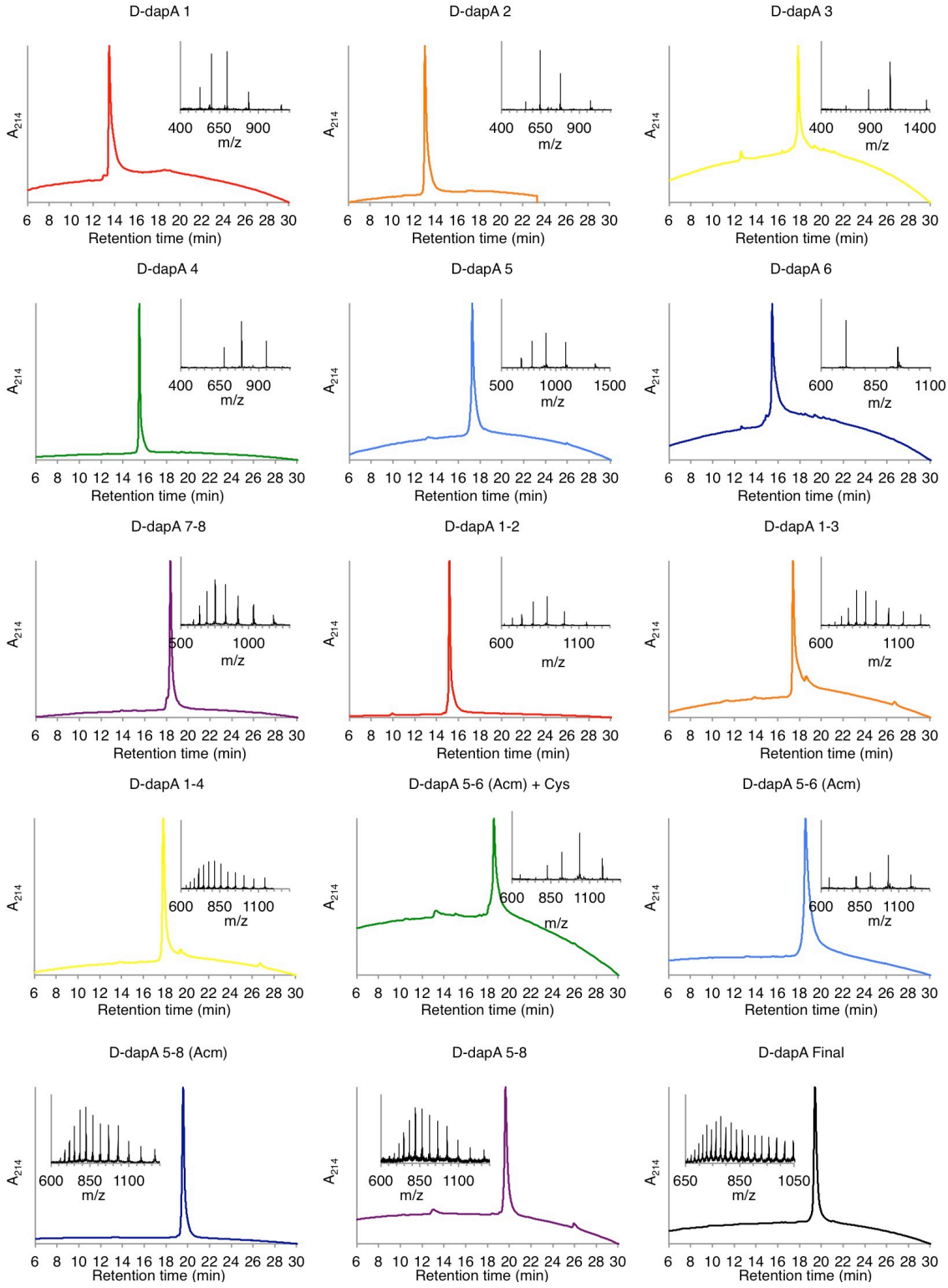


Figure 6.12. Characterization of assembly intermediates of D-dapA. Analytical RP-HPLC traces (C4 column) and mass spectra of individual peptides and ligation intermediates constructed for the synthesis of D-dapA A57C.

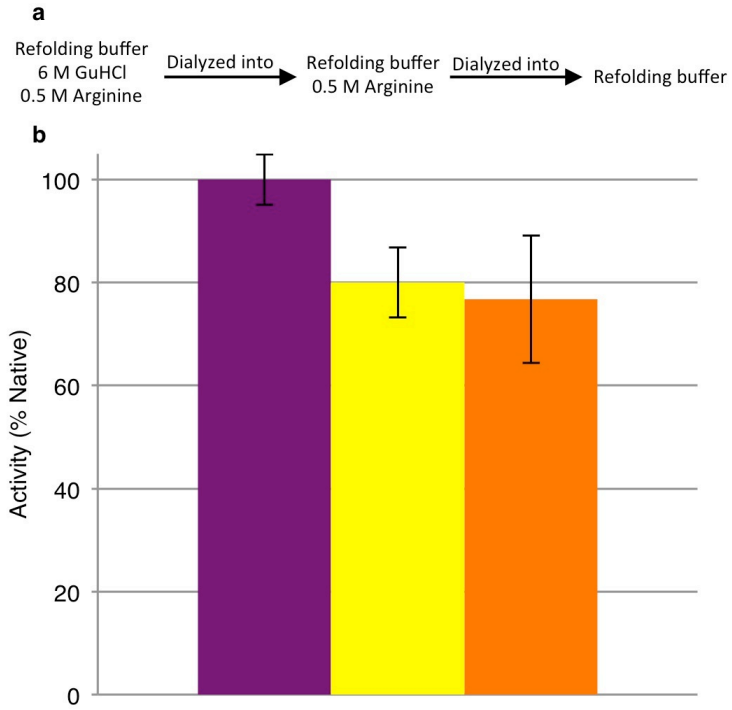


Figure 6.13. Arginine-assisted refolding of dapA. a, Schematic overview of arginine-assisted refolding protocol. **b**, Comparison of native recombinant dapA A57C activity (■) to recovered activity following arginine-assisted refolding with L-arginine (■) or D-arginine (■). Activity is normalized to native recombinant protein. Error bars indicate s.d. of at least three independent assays.

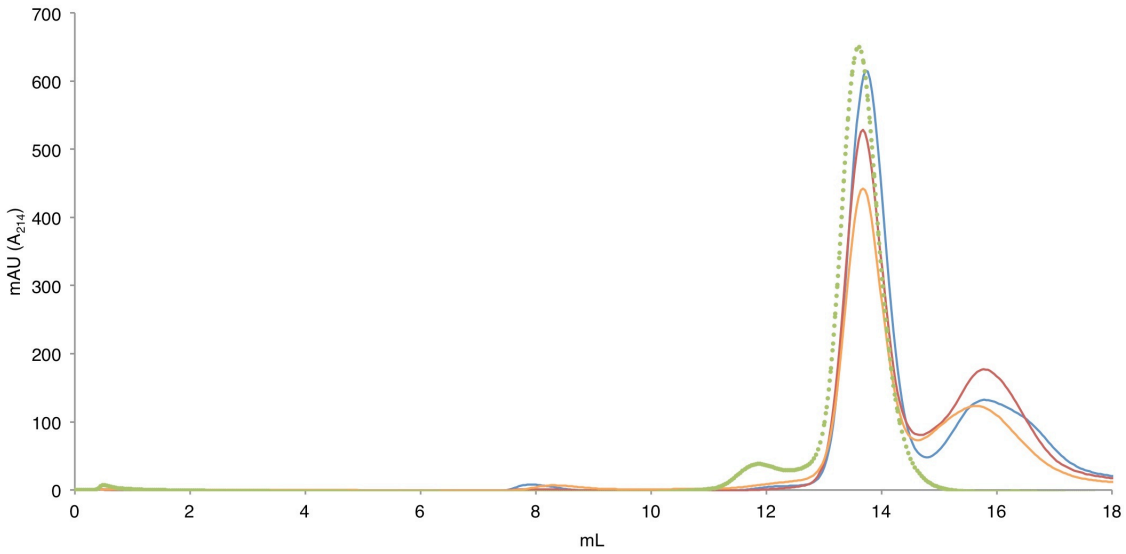


Figure 6.14. SEC purification of dapA following arginine-assisted refolding. Chromatograph from purification of arginine-assisted refolded synthetic L-dapA A57C (-) and D-dapA A57C (- and -) run on size-exclusion column (Superdex 200). Native recombinant dapA A57C tetramer (•) species (normalized) is shown for comparison.

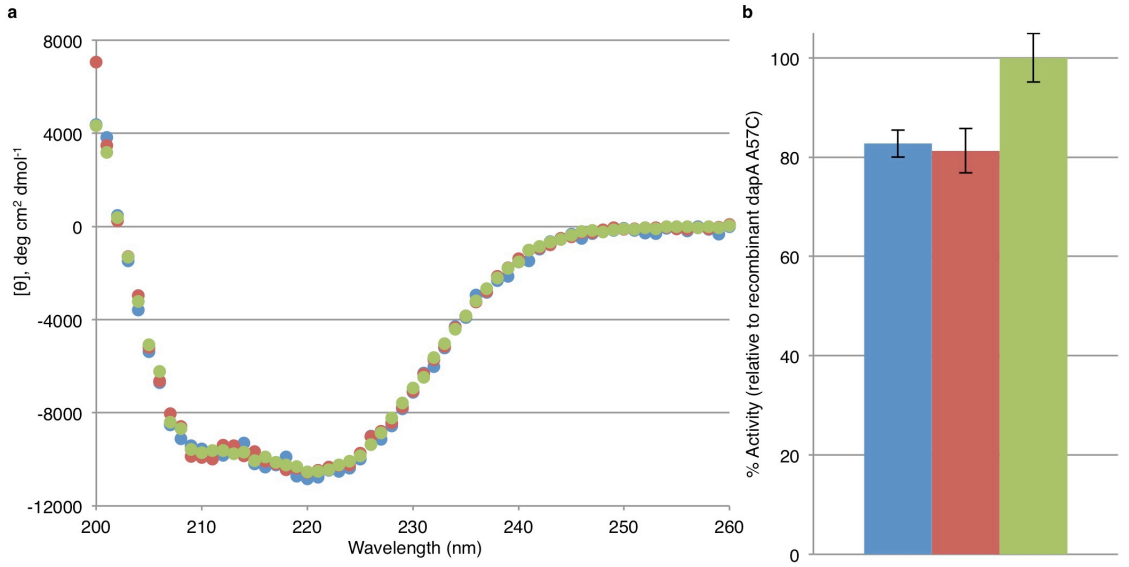


Figure 6.15. Structural and functional characterization of synthetic dapA constructs. a, Circular dichroism spectra of SEC-purified recombinant (●), synthetic L-dapA (●), and D-dapA (●) (the D-dapA spectrum has been multiplied by -1 to aid direct comparison with the other data). **b,** Activity of native recombinant dapA (■) and SEC-purified synthetic L-dapA (■) and D-dapA (■). Error bars indicate s.d. of at least three independent assays.

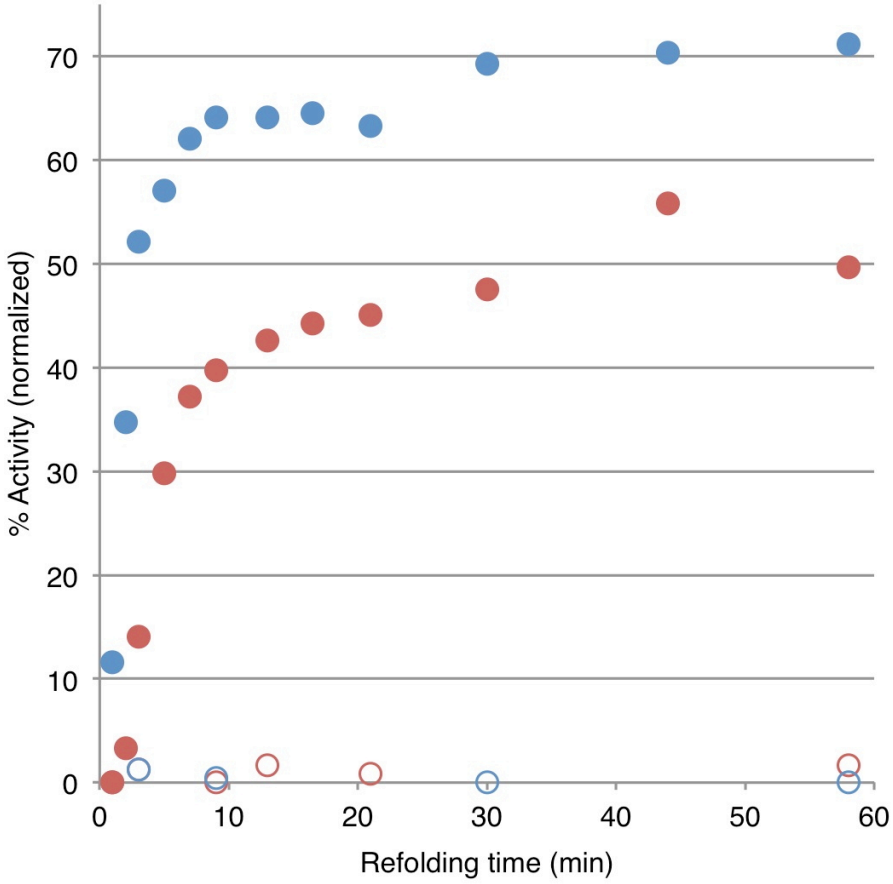


Figure 6.16. GroEL/ES-mediated refolding of synthetic dapA constructs. Refolding of synthetic L-dapA (●) and D-dapA (●) in the presence (closed circles) or absence (open circles) of GroEL/ES, normalized to the GroEL/ES-refolded recombinant dapA A57C activity level. Representative data shown from duplicate experiments.

References

1. Zawadzke LE, Berg JM. A racemic protein. Journal of the American Chemical Society. 1992;114(10):4002-3. doi: 10.1021/ja00036a073.

2. Forster AC, Church GM. Synthetic biology projects in vitro. Genome research. 2007;17(1):1-6. doi: 10.1101/gr.5776007. PubMed PMID: 17151344.

3. Horwich AL, Fenton WA. Chaperonin-mediated protein folding: using a central cavity to kinetically assist polypeptide chain folding. Quarterly reviews of biophysics. 2009;42(2):83-116. doi: 10.1017/S0033583509004764. PubMed PMID: 19638247.

4. Kim YE, Hipp MS, Bracher A, Hayer-Hartl M, Hartl FU. Molecular chaperone functions in protein folding and proteostasis. Annu Rev Biochem. 2013;82:323-55. doi: 10.1146/annurev-biochem-060208-092442. PubMed PMID: 23746257.

5. Houry WA, Frishman D, Eckerskorn C, Lottspeich F, Hartl FU. Identification of in vivo substrates of the chaperonin GroEL. Nature. 1999;402(6758):147-54. doi: 10.1038/45977. PubMed PMID: 10647006.

6. Kerner MJ, Naylor DJ, Ishihama Y, Maier T, Chang HC, Stines AP, et al. Proteome-wide analysis of chaperonin-dependent protein folding in Escherichia coli. Cell. 2005;122(2):209-20. doi: 10.1016/j.cell.2005.05.028. PubMed PMID: 16051146.

7. Fujiwara K, Ishihama Y, Nakahigashi K, Soga T, Taguchi H. A systematic survey of in vivo obligate chaperonin-dependent substrates. The EMBO journal. 2010;29(9):1552-64. doi: 10.1038/emboj.2010.52. PubMed PMID: 20360681; PubMed Central PMCID: PMC3212837.

8. Viitanen PV, Gatenby AA, Lorimer GH. Purified chaperonin 60 (groEL) interacts with the nonnative states of a multitude of Escherichia coli proteins. Protein Sci. 1992;1(3):363-9. doi: 10.1002/pro.5560010308. PubMed PMID: 1363913; PubMed Central PMCID: PMC2142211.

9. Lin Z, Schwartz FP, Eisenstein E. The hydrophobic nature of GroEL-substrate binding. J Biol Chem. 1995;270(3):1011-4. PubMed PMID: 7836352.

10. Fenton WA, Kashi Y, Furtak K, Horwich AL. Residues in chaperonin GroEL required for polypeptide binding and release. Nature. 1994;371(6498):614-9. doi: 10.1038/371614a0. PubMed PMID: 7935796.

11. Chen L, Sigler PB. The crystal structure of a GroEL/peptide complex: plasticity as a basis for substrate diversity. Cell. 1999;99(7):757-68. PubMed PMID: 10619429.

12. Wang Z, Feng H, Landry SJ, Maxwell J, Gierasch LM. Basis of substrate binding by the chaperonin GroEL. Biochemistry. 1999;38(39):12537-46. PubMed PMID: 10504222.

13. Coyle JE, Jaeger J, Gross M, Robinson CV, Radford SE. Structural and mechanistic consequences of polypeptide binding by GroEL. Folding & design. 1997;2(6):R93-104. PubMed PMID: 9427006.

14. Weinstock MT, Francis JN, Redman JS, Kay MS. Protease-resistant peptide design-empowering nature's fragile warriors against HIV. Biopolymers. 2012;98(5):431-42. doi: 10.1002/bip.22073. PubMed PMID: 23203688; PubMed Central PMCID: PMC3548907.

15. Welch BD, Francis JN, Redman JS, Paul S, Weinstock MT, Reeves JD, et al. Design of a potent D-peptide HIV-1 entry inhibitor with a strong barrier to resistance. J Virol. 2010;84(21):11235-44. doi: 10.1128/JVI.01339-10. PubMed PMID: 20719956; PubMed Central PMCID: PMC2953169.

16. Mirwaldt C, Korndorfer I, Huber R. The crystal structure of dihydrodipicolinate synthase from Escherichia coli at 2.5 A resolution. J Mol Biol. 1995;246(1):227-39. doi: 10.1006/jmbi.1994.0078. PubMed PMID: 7853400.

17. Blickling S, Renner C, Laber B, Pohlenz HD, Holak TA, Huber R. Reaction mechanism of Escherichia coli dihydrodipicolinate synthase investigated by X-ray crystallography and NMR spectroscopy. Biochemistry. 1997;36(1):24-33. doi: 10.1021/bi962272d. PubMed PMID: 8993314.

18. Devenish SR, Blunt JW, Gerrard JA. NMR studies uncover alternate substrates for dihydrodipicolinate synthase and suggest that dihydrodipicolinate reductase is also a dehydratase. J Med Chem. 2010;53(12):4808-12. doi: 10.1021/jm100349s. PubMed PMID: 20503968.

19. Yugari Y, Gilvarg C. The condensation step in diaminopimelate synthesis. J Biol Chem. 1965;240(12):4710-6. PubMed PMID: 5321309.

20. McLennan N, Masters M. GroE is vital for cell-wall synthesis. Nature. 1998;392(6672):139. doi: 10.1038/32317. PubMed PMID: 9515958.

21. Hackenberger CP, Schwarzer D. Chemoselective ligation and modification strategies for peptides and proteins. Angew Chem Int Ed Engl. 2008;47(52):10030-74. Epub 2008/12/17. doi: 10.1002/anie.200801313. PubMed PMID: 19072788.

22. Fang GM, Li YM, Shen F, Huang YC, Li JB, Lin Y, et al. Protein chemical synthesis by ligation of peptide hydrazides. Angew Chem Int Ed Engl. 2011;50(33):7645-9. doi: 10.1002/anie.201100996. PubMed PMID: 21648030.

23. Wan Q, Danishefsky SJ. Free-radical-based, specific desulfurization of cysteine: a powerful advance in the synthesis of polypeptides and glycopolypeptides. Angew Chem Int Ed Engl. 2007;46(48):9248-52. doi: 10.1002/anie.200704195. PubMed PMID: 18046687.

24. Mutter M, Nefzi A, Sato T, Sun X, Wahl F, Wohr T. Pseudo-prolines (psi Pro) for accessing "inaccessible" peptides. Pept Res. 1995;8(3):145-53. PubMed PMID: 7670229.

25. Wintermann F, Engelbrecht S. Reconstitution of the catalytic core of F-ATPase (alphabeta)3gamma from Escherichia coli using chemically synthesized subunit gamma. Angew Chem Int Ed Engl. 2013;52(4):1309-13. doi: 10.1002/anie.201206744. PubMed PMID: 23212931.

26. Tsumoto K, Umetsu M, Kumagai I, Ejima D, Philo JS, Arakawa T. Role of arginine in protein refolding, solubilization, and purification. Biotechnol Prog. 2004;20(5):1301-8. doi: 10.1021/bp0498793. PubMed PMID: 15458311.

27. Kohler RJ, Preuss M, Miller AD. Design of a molecular chaperone-assisted protein folding bioreactor. Biotechnol Prog. 2000;16(4):671-5. doi: 10.1021/bp0000609. PubMed PMID: 10933845.

28. Benoiton NL. Chemistry of Peptide Synthesis. Boca Raton, FL: CRC Press; 2006.

29. Mergler M, Dick F, Sax B, Weiler P, Vorherr T. The aspartimide problem in Fmoc-based SPPS. Part I. Journal of peptide science : an official publication of the European Peptide Society. 2003;9(1):36-46. doi: 10.1002/psc.430. PubMed PMID: 12587881.

30. Gibson DG, Glass JI, Lartigue C, Noskov VN, Chuang RY, Algire MA, et al. Creation of a bacterial cell controlled by a chemically synthesized genome. Science. 2010;329(5987):52-6. doi: 10.1126/science.1190719. PubMed PMID: 20488990.

31. Studier FW. Protein production by auto-induction in high density shaking cultures. Protein expression and purification. 2005;41(1):207-34. PubMed PMID: 15915565.

32. Kamireddi M, Eisenstein E, Reddy P. Stable expression and rapid purification of Escherichia coli GroEL and GroES chaperonins. Protein expression and purification. 1997;11(1):47-52. doi: 10.1006/prep.1997.0764. PubMed PMID: 9325138.

33. Haase C, Rohde H, Seitz O. Native chemical ligation at valine. Angew Chem Int Ed Engl. 2008;47(36):6807-10. doi: 10.1002/anie.200801590. PubMed PMID: 18626881.

34. Bang D, Chopra N, Kent SBH. Total chemical synthesis of crambin. Journal of the American Chemical Society. 2004;126(5):1377-83. doi: Doi 10.1021/Ja0385078. PubMed PMID: ISI:000188834900033.

35. Black S, Wright NG. Aspartic beta-semialdehyde dehydrogenase and aspartic beta-semialdehyde. J Biol Chem. 1955;213(1):39-50. PubMed PMID: 14353904.

36. Papadopoulos JS, Agarwala R. COBALT: constraint-based alignment tool for multiple protein sequences. Bioinformatics. 2007;23(9):1073-9. doi: 10.1093/bioinformatics/btm076. PubMed PMID: 17332019.

37. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics. 2009;25(9):1189-91. doi: 10.1093/bioinformatics/btp033. PubMed PMID: 19151095; PubMed Central PMCID: PMC2672624.