UNIVERSITY OF CALIFORNIA,
IRVINE

Accelerating directed evolution: self-mutating bacteriophage and controlled protein unfolding by shear-stress

DISSERTATION

submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in Biological Sciences

by

Tom Zhiye Yuan

Dissertation Committee:
Professor Gregory A. Weiss, Chair
Professor Celia Goulding
Professor Jennifer Prescher

2014

DEDICATION

To my parents and my brother,
whose love and support I will always grateful for.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

τ	Tau, shear stress
Ω	Omega, rotating tube velocity
v_θ	Fluid velocity
Å	Angstrom
ATP	Adenosine triphosphate
bAvd	*Bordetella* accessory variability determinant
BCA	Bicinchoninic acid assay
BP	*Bordetella bronchiseptica* bacteriophage
Brt	*Bordetella* reverse transcriptase
BSA	Bovine serum albumin
Bvg^+	Virulent *Bordetella bronchiseptica*
Bvg^-	Avirulent *Bordetella bronchiseptica*
cAMP	Cyclic adenosine monophosphate
CD	Circular dichroism
cDNA	Complementary DNA
CDR	Complementarity-determining region
C_H	Constant heavy domain
C_L	Constant light domain
CLec	C-type lectin
DARPin	Designed ankryin-repeat repeat protein
DGR	Diversity-generating retroelement
dsDNA	Double-stranded DNA

ELISA	Enzyme-linked immunosorbent assay
FDA	Food and Drug Administration
FDC	Follicular dendritic cells
Fab	Fragment-antigen binding
gp41	Glycoprotein 41
GPCR	G protein-coupled receptor
GST	Glutathione s-transferase
HEWL	Hen egg white lysozyme
HIC	Hydrophobic interaction chromatography
HIV	Human immunodeficiency virus
HRP	Horseradish peroxidase
Ig	Immunoglobulin
IMAC	Immobilized metal ion affinity chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
K_D	Dissociation constant
LDH	Lactate dehydrogenase
L-Mtd	Lysozyme-binding major tropism determinant
LRR	Leucine-rich repeat
MBP	Maltose binding protein
Mtd	Major tropism determinant
NADH	Nicotinamide adenine dinucleotide
NFM	Non-fat milk
NusA	N-utilizing substance A

OPD	*o*-Phenylenediamine
Pa	Pascal
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PDB	Protein data bank
PFU	Plaque forming unit
PKA	Protein kinase A
PSA	Prostate specific antigen
PSMA	Prostate specific membrane antigen
rhGH	Recombinant human growth hormone
rpm	Rotations per minute
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEC	Size-exclusion chromatography
scFv	Single-chain variable fragment
SMPL	Self-made phage library
SPRI	Surface plasmon resonance imaging
SUMO	Small ubiquitin-like modifier
TR	Template region
VFD	Vortex fluid device
V_H	Variable heavy domain
V_L	Variable light domain
VLR	Variable lymphocyte receptor
VR	Variable region

ACKNOWLEDGEMENTS

I would like to thank my advisor, Professor Gregory A. Weiss, for his endless enthusiasm and unwavering support for his students. He has been an irreplaceable source of encouragement, patience, and inspiration.

I would like to thank Dr. Cathie M. Overstreet, who trained me from day one in the laboratory, and has always been supportive throughout my graduate school journey even after she had moved onto her true passion – teaching.

I would like to thank all my fellow Weiss lab members – in particular Dr. Sudipta Majumdar and Dr. Jessica Arter, for inspiring me; Dr. Issa Moody, who never failed to entertain; Kritika Mohan and Stephan Kudlacek, for being amazing partners in various projects.

I would like to acknowledge Professor Jennifer Prescher and Professor Celia Goulding for serving on my committee and providing valuable feedback.

Last, but certainly not least, I would like to thank my family for their encouragement and patience. I would also like to thank my friends, who are still unable to comprehend what I'm talking about, or how I'm employable.

Tom Zhiye Yuan
Curriculum Vitae

254 Ventura Avenue
Palo Alto, CA 94306
Tel: (650) 776-5818
Email: tzyuan@uci.edu

Education

July 2014
University of California, Irvine Doctor of Philosophy, Biological Sciences

June 2007
University of California, Davis Bachelor of Science, Biochemistry & Molecular Biology

Research Experience

August 2008 – July 2014
University of California, Irvine
Department of Molecular Biology and Biochemistry
Irvine, CA
Graduate Student Researcher
Study and characterization of a unique in vitro molecular evolution method, analogous to antibody maturation, utilizing *Bordetella* phage display to identify high affinity non-immunoglobulin protein ligands. Directed evolution of M13 phage-displayed peptide and protein libraries targeting prostate and bladder cancer biomarkers for use in real-time, organic biosensors. Investigated shear stress applications to solubilize misfolded and aggregated protein from inclusion bodies.

June 2007 – August 2008
Genencor, DuPont Industrial BioSciences
Palo Alto, CA
Research Assistant
Assay development and full-length protein engineering of protease targets for use in detergent applications. Subtilisin BPN' protease screen conducted in collaboration with research and development branch in Leiden, Netherlands.

Publications

Yuan, T.Z., Overstreet, C.M., Moody, I.S., Weiss, G.A. (2013). Protein Engineering with Biosynthesized Libraries from Bordetella bronchiseptica Bacteriophage. *PLoS One*. 8: e55617.

Overstreet, C.M., **Yuan, T.Z.**, Levin, A.M., Kong, C., Coroneus, J.G., Weiss, G.A. (2012). Self-made phage libraries with heterologous inserts in the Mtd of Bordetella bronchiseptica. *PEDS*. 25: 145-151.

Arter, J.A., Diaz, J.E., Donavan, K.C., **Yuan, T.**, Penner, R.M., Weiss, G.A. (2012). Virus–Polymer Hybrid Nanowires Tailored to Detect Prostate-Specific Membrane Antigen. *Anal. Chem.* 84: 2776-2783.

Nieto, N.C., Foley, J.E., MacLachlan, N.J., **Yuan, T.Z.**, Spier, S.J. (2009). Evaluation of hepatic disease in mice following intradermal inoculation with Corynebacterium pseudotuberculosis. *Am. J. Vet. Res.* 70: 257-262.

Patents

Weiss, G.A., Raston, C.L., **Yuan, T.Z.**, Ormonde, C. 2013. Method for protein refolding using vortexing fluid shear forces. U.S. Provisional Serial 61/873,718, filed September 2013. *Patent Pending*.

Presentations

Yuan, T.Z., Ormonde, C., Raston, C.L., Weiss, G.A. (2013). Utilizing fluid shear forces as a mechanical chaperone for protein folding. *Molecular Biology & Biochemistry Department Retreat*. Lake Arrowhead, CA. [Oral presentation]

Yuan, T.Z., Overstreet, C.M., Levin, A.M., Moody, I.S., Kong, C., Coroneus, J.G., Weiss, G.A. (2012). Self-made phage libraries optimized for target binding. *243st National Meeting of the American Chemical Society*. San Diego, CA. [Poster]

Yuan, T.Z., Overstreet, C.M. (2010). Protein engineering with biosynthesized libraries from Bordetella bronchiseptica bacteriophage. *Future Diagnostics Conference*. Irvine, CA. [Poster]

Yuan, T.Z., Overstreet, C.M., Levin, A.M., Weiss, G.A. (2010). Protein engineering with biosynthesized libraries from Bordetella bronchiseptica bacteriophage. *239th National Meeting of the American Chemical Society*. San Francisco, CA. [Poster]

Laboratory Techniques

Phage display, directed evolution, protein and peptide library design and selection.

Site-specific protein modification, recombinant protein expression and purification (FPLC, HPLC, IEC, SEC, IMAC, GST, MBP).

Biophysical assay development, molecular biology, and protein chemistry (SPR, ITC, MALDI-TOF, CD, DLS, UV/Vis, ELISA).

Proficiency in GraphPad Prism, SoftMax Pro, LaserGene, Vector NTI, SAP, MS Office, Linux, and Mandarin Chinese.

Teaching and Mentoring

August 2010 – June 2014
University of California, Irvine
Department of Molecular Biology and Biochemistry
Graduate Student Mentor
Mentored incoming graduate and undergraduate students on projects spanning protein purification, library design and selection, and bacteriophage display technologies.

August 2010 – August 2014
University of California, Irvine
Department of Molecular Biology and Biochemistry
Teaching Assistant
Participated in planning, design and execution of microbiology laboratory sections. Individually taught molecular biology and biochemistry discussion sections.

June 2011 – May 2012
American Association for the Advancement of Science
Irvine, CA and Mohammedia, Greater Casablanca, Morocco
Training in Nanobiotechnology for Detection of Environmental Viruses
Long-term collaboration between Moroccan and American investigators to develop novel methods for identifying and diagnosing environmental pathogens.

ABSTRACT OF THE DISSERTATION

Accelerating directed evolution: self-mutating bacteriophage and controlled protein unfolding by shear-stress

By

Tom Zhiye Yuan

Doctor of Philosophy in Biological Sciences

University of California, Irvine, 2014

Professor Gregory A. Weiss, Chair

Proteins, large biological molecules synthesized by living organisms, serve a wide array of functions. Many proteins operate as molecular scaffolds for binding to other proteins; in fact, the vast majority of molecular interactions in biology are made possible by protein-protein interactions. Over the past 15 years, powerful techniques have been developed to generate protein-based ligands *in vitro* to virtually any protein target. The fundamental steps to any protein engineering effort remain the same. The protein target of interest is modified to create a mutant with preferable biochemical characteristics. The protein is then expressed and purified in large quantities for use in therapeutic or diagnostic applications. In this thesis, I describe a multifaceted approach to address the diversification, identification, and production steps of the protein engineering process.

Currently, the vast majority of binding molecules that have been developed for use in biomedical research are either antibodies or antibody derivatives.

Immunoglobulins, such as antibodies, are notoriously difficult to produce. These large, multi-domain proteins require complex cloning steps for recombinant expression in mammalian cell lines. Alternatively, researchers have begun to engineer binding proteins outside the immunoglobulin family by using protein scaffolds with structurally rigid cores accompanied by solvent-accessible surface loops. We characterize such a protein scaffold, the major tropism determinant of *Bordetella bronchiseptica* bacteriophage, which is under the control of a diversity generating genetic element. This approach allows automated generation of a large phage-displayed protein receptor library for use in selection experiments to identify binders to any protein target of choice.

Following the identification of the protein ligand, successful recombinant expression of the protein is crucial to commercial viability. In bacterial expression hosts, proteins will often misfold and clump into inclusion bodies. This protein aggregation results in drastically decreased expression yields. Returning the misfolded protein to the native state conformation by traditional methods requires laborious and time-consuming processing steps. We investigate a novel method for applying shear stress to rapidly refold proteins from inclusion bodies. This research, combined with further studies of non-immunoglobulin protein scaffolds, could drastically lower the cost of protein engineering efforts and enable new biological applications in the future.

CHAPTER 1

Accelerating directed evolution

Abstract

A directed evolution approach to protein engineering allows investigators to bypass typically lengthy traditional rational design steps by directly enriching for desired characteristics of the target protein. Typically, these methods require no intimate knowledge of the protein structure or enzyme mechanism. Directed evolution mimics the natural selection process that evolves proteins, but the techniques applied operate on significantly shorter timescales. For the purpose of discovering new protein-protein binding motifs, nature suggests many robust structural motifs, including immunoglobulins and carbohydrate binding c-type lectins. These two classes of proteins can be expressed under the control of diversity generating genetic elements, described in this chapter, and are subject to selection pressures designed to quickly identify ligands to novel antigens or protein receptors. Following successful identification of a candidate ligand, experimenters are often challenged with the recombinant expression of a stubbornly insoluble protein. Once again, nature provides an excellent example of overcoming this challenge through chaperone proteins. The GroEL-GroES complex in *Escherichia coli*, for example, forcibly unfolds misfolded proteins, giving the misshapen polypeptides a second chance for reconfiguration. Recent advances in the application of mechanical forces, such as high pressures or shear stress, show promising strides towards the controlled unfolding of protein aggregates. In a similar fashion to the formation of entirely new industries from advances in high-throughput sequencing, steady advances in protein engineering and expression methods could lower the cost of

entry and enable new biological applications to control cell biology.

Introduction

Protein engineering improves or redesigns the function of an existing protein for a variety of applications – including but not limited to gene expression, cellular gatekeeping, chemical catalysis and inhibition[1]. Protein macromolecules are comprised of linear amino acid polymers connected by peptide bonds[2]. Amino acid residues display 20 naturally occurring side chains, each imparting specific chemical properties within the context of the local microenvironment. Previous efforts to improve or modify naturally evolved proteins have generated notable industrial, therapeutic, and diagnostic successes. Examples include an engineered protease that retains activity at a wide range of temperatures and in the presence of bleach[3], a long-lasting formulation of insulin[4], and an antibody fragment that binds to the hormone testosterone[5].

Proteins have steadily become a major class of therapeutics and diagnostics since protein engineering techniques were pioneered in the early 1980s[6]. Over 200 biopharmaceuticals have been approved in the United States and the European Union[7]. Protein engineering has enabled investigators to enhance desirable characteristics such as high affinity, specificity, and low immunogenicity, while also increasing our understanding of protein-protein binding interfaces[8–10]. However, established methods to design and validate even a single target represent a major undertaking for academic and commercial labs alike. Current approaches require the production of large libraries by site-directed mutagenesis or other means of diversification. Presently, the most common protein engineered to bind novel targets are the vertebrate immunoglobulins, or antibodies[11]. Unfortunately these large multi-domain, glycosylated proteins remain problematic due to high dosage requirements and low expression yields[12]. The work in

this thesis addresses the two bottlenecks described above. First, we repurpose a novel self-mutating protein ligand found in *Bordetella bronchiseptica* bacteriophage to bind to a designated protein target, without the use of site-directed mutagenesis. To lower the cost of production, we developed a rapid, low-cost method using controlled shear stress to recover misfolded protein and boost low expression yields.

Strategies for protein engineering are defined by a spectrum of methods that require either a vast library diversity, termed directed evolution or intimate knowledge of the target protein structure and enzyme mechanism, termed rational design. If the protein target has already been exhaustively studied, typically including a three-dimensional structure determined by NMR spectroscopy or X-ray crystallography, a rational design approach allows experiments to focus amino acid substitutions on a defined region of protein sequence space. Applying this approach limits the subsequent screening effort to a manageable number of variants. Both directed evolution and rational design approaches have proven successful in previous experiments. For example, directed evolution methods identified aldolase variants with altered substrate specificities[13]; rational design experiments yielded mutants with altered stereoselectivity[14] and enhanced stability[15].

While both directed evolution and rational design strategies both involve site-directed mutagenesis, directed evolution methods hold the distinct advantage of allowing investigators to skip the time-consuming rational design process entirely[16]. The technique repurposes the natural selection process with gene mutation and selection for characteristics defined as advantageous by the experimentalist. This *in vitro* method can be performed in considerably shorter timescales relative to natural selection in

nature. At first glance, artificial efforts to improve protein functionality appear daunting. Natural selection has occurred over billions of years, and even small proteins could be combinatorialized into an astronomical number of potential sequence variations (>10^{260} variants for a 200 residue protein). Most directed evolution efforts are conducted in 96-well microtiter plates and typically screen less than 2000 clones[17]. Despite these challenges, numerous directed evolution studies have improved protein characteristics, including binding affinity, enzyme turnover, substrate specificity, and protein stability without detailed structural and mechanistic information[18–22]. Reviews published by Farinas *et al*.[23], and Tao *et al.*[19] summarize efforts to elucidate how evolutionary approaches can improve protein structure and function.

As with any laboratory technique, each step of the directed evolution process requires optimization. Maximizing library diversity is vital to effective directed evolution experiment. Following library generation, selection techniques must fine-tune stringency to avoid premature removal of potential binding partners. Additionally, amplification-induced biases should be minimized during each round of library propagation. Lastly, expression and purification of the protein candidate must be economically viable (Figure 1-1).

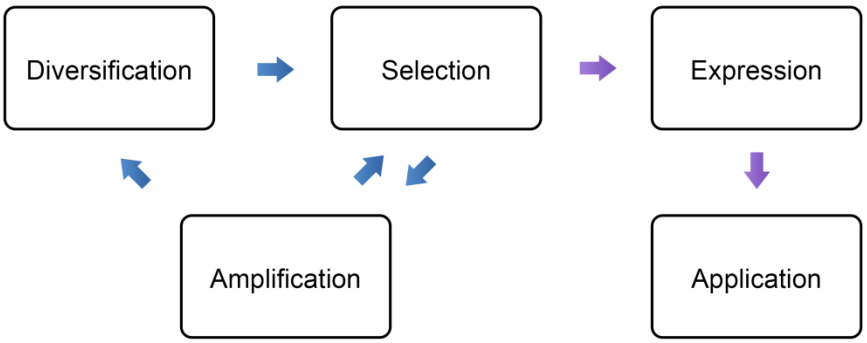


Figure 1-1. General scheme for directed evolution. The initial level of diversity is critical to conducting a successful directed evolution experiment. Next, selections screen for variants that improve protein properties and functionality. Multiple selection and amplification steps enrich the protein library before maturation by further diversification. Following the identification of a lead candidate, the engineered protein is eventually expressed and used in downstream applications.

1.1 Naturally evolved protein-protein binding motifs

Protein-protein interactions are vital components for the vast majority of molecular interactions in biology[24,25]. Protein-protein binding interfaces rely largely on shape complementarity and non-covalent forces such as hydrogen bonding, electrostatic interactions, van der Waals forces, and π-π or cation-π interactions[26]. As these forces are only effective at limited distances, binding interfaces are typically very closely aligned and the total binding energy is diffused over a large area (up to 1600 Å^2)[25,27,28].

Nature provides numerous examples of proteins ligands that have evolved to bind other proteins. Investigation of natural protein-protein interactions can guide strategies to identify novel receptor-ligand pairs. The benchmark example is the antibody (immunoglobulin, Ig) protein expressed by jawed vertebrates. These proteins possess a rigid, tertiary structure able to tolerate massive sequence variation, and enable the organism to access the widest possible array of receptor-ligand pairs[29].

Antibody-antigen binding sites are typically comprised of large residues capable of participating in electrostatic and van der Waals interactions, interspersed with small, flexible amino acids which allow structural flexibility. These sites also tend to favor amphipathic amino acids such as tyrosine and tryptophan that are able to tolerate both hydrophilic (exposed) and hydrophobic (buried) environments[29,30].

1.1.1 Diversity generation in antibodies

The vertebrate immune system uses the immunoglobulin scaffold to respond to foreign molecules and organisms by introducing massive sequence variation to well-defined hypervariable loop regions. The successful production of antibody–antigen pairs require the enrichment and eventual selection of a specific, high-affinity Ig-based ligand from an initial library of $\approx 10^{14} - 10^{16}$ immunoglobulin variants[31,32]. B lymphocytes, or B cells, achieve this level of diversity by a combination of V(D)J recombination, somatic hypermutation, and affinity maturation. In V(D)J recombination, immature B cells randomly select one variable (V), one diversity (D), and one joining (J) gene out of multiple copies of the same gene segment[33]. During somatic hypermutation, mutations in the complementarity-determining region (CDR) average one nucleotide change per gene per cell division[34]. Following mutagenesis, B cells compete for a limited amount of antigen presented by follicular dendritic cells (FDC) and B cells variants that display high binding affinity to presented antigen are selected for survival[34]. Repeated selection rounds enrich for antibodies with the highest affinity binding for the antigen.

1.1.2 Non-immunoglobulin protein ligands in nature

The adaptive immune response developed by jawed vertebrates is not the sole example of diversity generation in nature for producing binding protein ligands. Jawless vertebrate sea lampreys, for example, selectively express a series of leucine-rich repeat (LRR) genes to produce dynamic libraries of the variable lymphocyte receptor (VLR)[35]. These VLR proteins function as the primary antigen recognition receptor in the sea lamprey adaptive immune system[36]. In addition to high primary sequence diversity, the number of LRR repeats can vary from variant to variant. The flexibility in protein size adds another dimension to structural diversity by allowing the receptor to adapt to larger or smaller binding targets. The modular architecture of repeat proteins allows easy expansion of library diversity and constitutes a significant share of characterized protein-protein interactions found in nature.

1.1.3 Artificially engineered protein-protein binding scaffolds

Other structural motifs are also found throughout nature to mediate protein-protein interactions, but are not under the control of diversity generating genetic elements. Instead, experimenters have created artificial protein libraries *in vitro* based on common protein-protein binding motifs (Figure 1-2). The ankryin-repeat is a common motif found in eukaryotes, prokaryotes, and archea that has been adapted for use in designed ankryin-repeat protein (DARPin) libraries. The libraries were designed by connecting multiple ankryin-repeat motifs and randomizing residues as guided by analysis of the sequence and structural consensus[37]. These libraries were expressed by ribosome-display and selections identified ligands to maltose binding protein[37,38]. Further experiments have demonstrated that the DARPin libraries are capable of

producing inhibitors to aminoglycoside-phosphotransferases[39]. DARPins also can provide high expression yields in bacterial hosts, partly due to a lack of cysteine residues[37]. Wikman and colleagues report utilizing the protein A-derived Z domain in affibody protein libraries to identify high-affinity receptors for the breast cancer biomarker human epidermal growth factor receptor II[40].

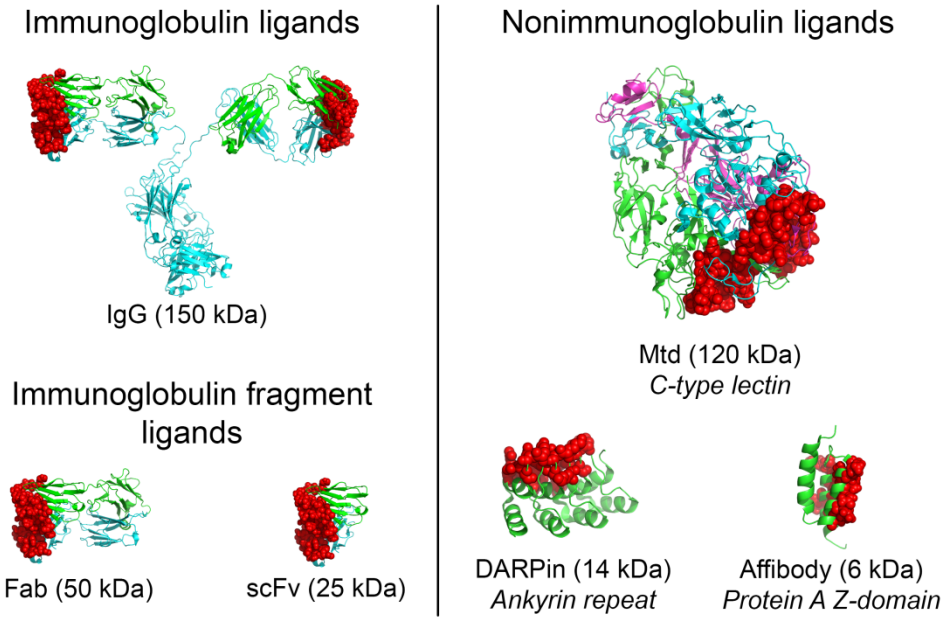


Figure 1-2. Overview of immunoglobulin and non-immunoglobulin based protein receptors. On the left, immunoglobulin G (PDB: 1IGT)[41] is shown as the prototypical antibody along with the most commonly developed antibody fragments – fragment-antigen binding (Fab) fragments and single-chain variable fragments (scFv). Heavy and light chains are shown as cyan and green ribbons, respectively. Residues corresponding to the mutagenic complementary determining region are displayed as red spheres The right shows ribbon representations of the non-immunoglobulin protein ligands major tropism determinant (Mtd, PDB: 1YU0)[42], designed ankryin-repeat protein (DARPIN, PDB: 1N0R)[43], and affibody (PDB: 2B89)[44], listed with their constituent structural motifs. Residues targeted for diversification in directed evolution studies are displayed as red spheres.

1.1.4 Directed evolution approaches to identify protein ligands

Approaches to the design and selection of large protein libraries *in vitro* can apply either immunoglobulin or non-immunoglobulin protein domains. Whole antibodies are rarely employed for library generation due to expression difficulties stemming from their extensive glycosylation, disulfide bond formation and multi-domain architectures. Smaller fragments of the antibody are also better suited to recombinant expression in bacterial and yeast systems (Figure 1-2). Directed evolution strategies utilizing recombinant fragment antigen-binding (Fab) and single-chain variable fragment (scFv) libraries have resulted in therapeutics approved by US Food and Drug Administration (FDA) approval, including ranibizumab (Genencor) for macular degeneration and efungumab (Novartis) for invasive *Candida* infection[45,46].

Previous selection studies by Söderlind *et al*.[47] and Holt *et al.*[48] demonstrate that novel binding proteins can be readily identified from Fab and scFV libraries. Unfortunately, the stability of antibody fragment proteins is typically reliant on intradomain disulfide bonds, which cannot form in the reducing cytosolic space of bacterial host[49]. The convoluted patent landscape for antibody technologies further complicates antibody engineering efforts[6]. With a more readily expressed non-immunoglobulin protein scaffold, investigators can reduce the challenges involved in the expression of a multidomain, glycosylated eukaryotic proteins. To take advantage of the efficiencies inherent to a non-immunoglobulin scaffold, our laboratory has focused efforts on a bacteriophage protein that has evolved into a generalized protein receptor.

1.2 Major tropism determinant as a novel protein ligand scaffold

The ideal protein scaffold would be amneable to overexpression in a bacterial host, not require or include disulfide bonds, and tolerate high levels of diversity without compromising structural stability. One such protein is the major tropism determinant (Mtd) of *Bordetella bronchiseptica* bacteriophage. A receptor binding protein that has adopted a common carbohydrate binding motif, the C-type lectin fold (CLec) into a general ligand-binding structure (Figure 1-3). Variants of the Mtd possess nearly identical structures thus demonstrating its functionality as a rigid scaffold[42,50]. The combination of massive sequence variation and high structural stability is crucial to identifying protein ligands by directed evolution.

The challenge for any directed evolution project is the sheer number of possible permutations, resulting from combinatorial amino acid substitutions. The generation of a very large protein library is non-trivial, and can require a major effort depending on the size of the protein. Error-prone PCR can mutate, on average, the DNA encoding approximately only one in every six amino acids, thus making mutations in adjacent residues unlikely[51]. Utilizing the Mtd as a protein scaffold addresses one of the major challenges of library generation. The gene encoding for the Mtd undergoes massive sequence variation during each round of phage propagation.

1.2.1 The diversity-generating retroelement

In 2002, Liu *et al.* identified the genetic cassette that diversifies the gene encoding the Mtd, termed a diversity-generating retroelement (DGR)[52]. DGRs are the first prokaryotic genetic element discovered that generate massive sequence variability for the purpose of identifying novel protein ligands[52,53].

In *Bordetella* bacteriophage, a T7-like dsDNA phage, the DGR mutates the *mtd* gene that determines host tropism of the virus particle. Switching tropism is advantageous to phage survival due to the rapidly-changing surface receptor proteins expressed on its bacterial host, *B. bronchiseptica.* This bacterium switches the profile of its surface-displayed proteins during transitions between virulent and avirulent phases[54,55]. Diversification is facilitated by a phage-encoded reverse transcriptase that substitutes any adenine coded amino acid within a non-coding template region (TR). This allows the bacteriophage to generate up to 10^{13} unique Mtd variants, thus providing an extensive collection of potential receptor-ligand pairs[50,52,56].

1.2.2 Major tropism determinant library generation

The *mtd* gene encoding the eponymously named protein contains a 134 bp sequence near the 3' end that corresponds to substitutions at 23 distinct positions near the C-terminus of the Mtd. This is termed the variable region, or VR. A non-coding copy of the 134 bp sequence designated as the template region, or TR, is located downstream from the *mtd*. In contrast to the VR, the TR remains invariant from phage to phage. The retroelement also encodes for *Bordetella* reverse transcriptase (Brt) that diversifies only adenosine base pairs. Sequence information from the non-coding template region of the genome replaces the coding variable region of the gene after the Brt replaces adenines with any of the four DNA bases[52,56].

1.3 The major tropism determinant as a protein scaffold

The Mtd protein is displayed as a homotrimer at the end of the *Bordetella* phage tail fibers as the main receptor-binding ligand. The Mtd contains a unique C-type lectin (CLec) fold at the C-terminus and organizes a relatively flat receptor-binding surface (Figure 1-3). Trimerization is primarily mediated by a large hydrophobic surface (>4500 Å) at the N-terminus of Mtd[42]. Each monomer is comprised of an N-terminal domain defined by β-prism-forming sheets, an intermediate domain dominated by a β-sandwich made of antiparallel sheets, and a C-terminal domain with the CLec fold. The mutable VR region of the Mtd is located in this CLec fold. Though structurally similar to macrophage mannose receptor and intimin, both calcium-dependent carbohydrate binding proteins, the CLec fold employed by *Bordetella* bacteriophage lack the corresponding calcium and carbohydrate binding residues[57–60]. The Mtd CLec further differentiates itself from other related variants by the addition of a short β4' strand at the C-terminus. The Mtd CLec also lacks the four disulfide-bonded cysteines typical of most variants employed in animal lectins, thus simplifying the bacterial expression of recombinant Mtd[42].

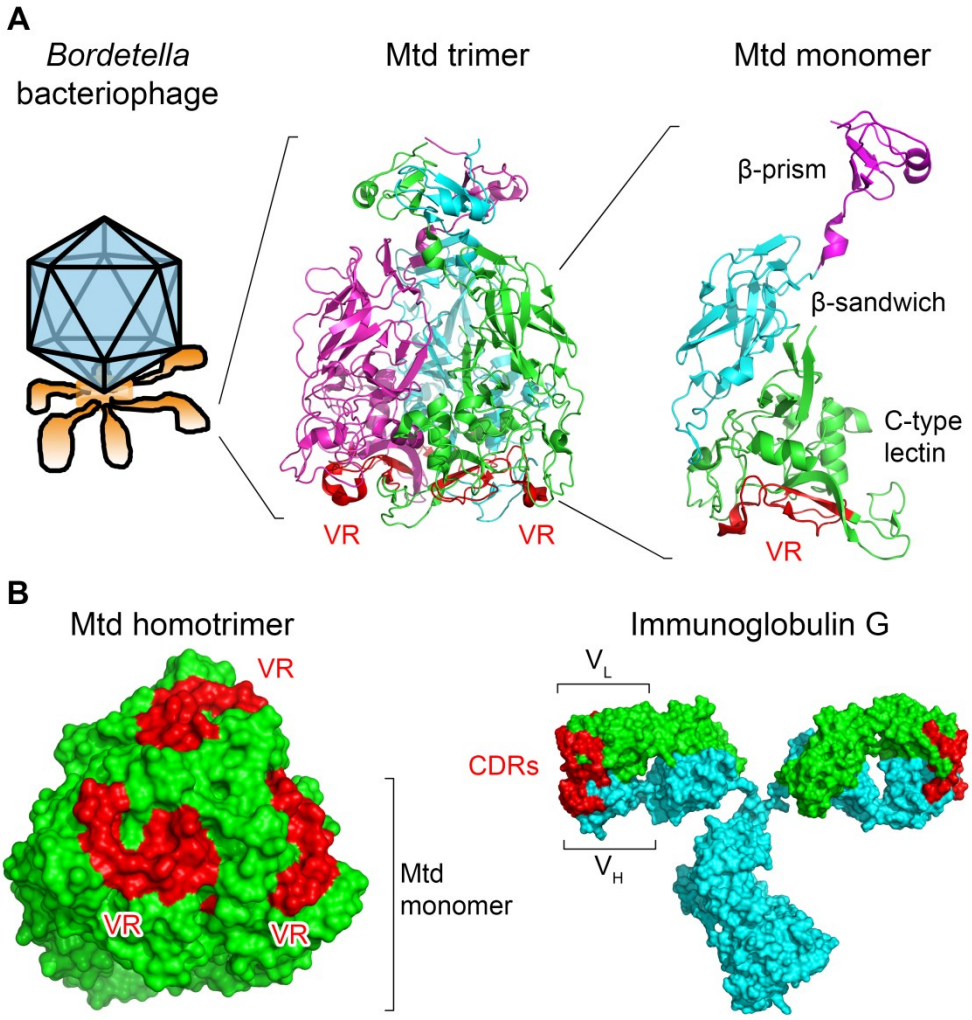


Figure 1-3. Major tropism determinant functions as the receptor-binding protein of *Bordetella* bacteriophage. (A) The bacteriophage displays six tail fibers, each containing two homotrimers of the Mtd protein[61]. The mutagenic variable regions (VR) (red) are organized in a relatively flat receptor-binding surface (PDB: 1YU4)[61]. (B) Surface representations, with regions targeted for subsitutions by mutagenesis (red), of the Mtd homotrimer and immunoglobulin G (PDB: 1IGT)[41].

The *Bordetella* phage takes advantage of the homotrimerization of Mtd to display three identical VR sequences in close proximity at the receptor-binding surface[61]. In contrast, the immunoglobulin G displays six non-identical CDRs on each of side of the Y-shaped protein (Figure 1-3). Though antibodies have an advantage in diversity levels

(≈10^{14} – 10^{16})[31,32], Mtd libraries approach these diversities with ≈10^{13} possible sequences, and also avoid bacterial expression and patent limitations[42,56]

1.4 Adapting diversity-generating retroelements for *in vitro* applications

Bordetella phage allows experimenters to bypass the time-consuming mutagenesis step inherent to directed evolution by providing a self-mutating library. The Mtd a scaffold has already evolved for general protein-binding activity. Utilizing the Mtd allows researchers to skip the non-trivial step of designing, generating, and cloning peptide and protein libraries. However, selections with *Bordetella* phage require a radical departure from typical strategies utilized in previously developed phage display methodologies (Figure 1-4). The most common phage display methods utilize filamentous M13 bacteriophage, first described in 1985 by George Smith[62]. In most phage display selections, the initial peptide or protein library is generated prior to multiple rounds selection and enrichment without further diversification between rounds, allowing optimization of selection stringencies and rotating selection strategies between rounds. M13 phage display of immunoglobulin fragments such as fragment antigen-binding (Fab fragment)[63–65] and single-chain variable fragment (scFv)[66–68] libraries previously identified high-affinity ligands to protein, peptide, and nucleotide targets.

In *Bordetella* phage display, multiple selection rounds are counter-productive as the enrichment step would be nullified by Brt-mediated mutagenesis upon phage propagation. Instead, optimization is focused on the initial diversification process to allow maximum library diversity with multiple rounds of diversification and amplification. Next, a single round of selection is performed – highly unusual among phage-based

selection techniques (Figure 1-4). The approach has until our reports, not been explored[69].

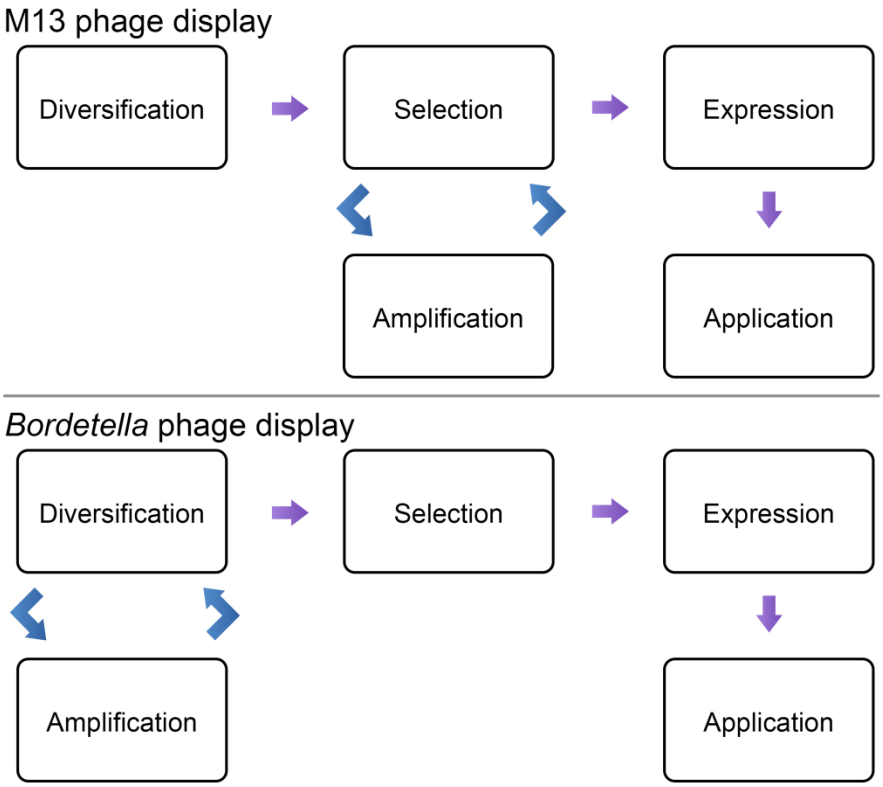


Figure 1-4. M13 phage display vs *Bordetella* phage display strategies. M13 phage display typically generates all library diversity before the first selection step. Conversely, *Bordetella* phage display undergoes multiple diversification and amplification rounds before a single selection step.

1.5 Further considerations in bacterial expression systems

While utilizing Mtd and similar scaffolds that are amenable to high-yield, recombinant expression in bacterial hosts can reduce downstream production costs, prioritizing protein targets by favorable expression characteristics limits investigators from exploring vast areas of protein diversity space. Protein solubility often remains unpredictable. Expressing eukaryotic and large, multi-domain proteins in bacterial vectors are highly susceptible to protein aggregation into inclusion bodies (Table 1-1)[70–

[72]. Experiments conducted by Wang *et al.* suggest that bacterial inclusion bodies are not typically amorphous aggregates as previously believed, but instead are highly structured as amyloid-like fibrils, which are rich in β-sheets[73].

	Advantages	Disadvantages
Bacterial	High yield	Lack of post-translational modifications
	Lowest cost	Eukaryotic proteins typically insoluble
	Rapid production	Reducing environment interferes with disulfide bond formation
	Most common setup	Membrane proteins difficult
Yeast	Relatively simple	Incomplete post-translational modifications
	Lower cost	Membrane proteins difficult
Insect	Post-translational modifications	Slower production
		Higher cost
		Membrane proteins difficult
Mammalian	Native expression environment	Slowest production
	Post-translational modifications	Highest cost
		Low yield

Table 1-1. Advantages and disadvantageous of various recombinant protein expression hosts[74,75].

The rapid rate of protein synthesis in *E. coli* is advantageous from a recombinant expression viewpoint but promotes self-association of the nascent polypeptide chains, which leads to aggregation[76]. Expression temperatures can be lowered to slow the rate of protein synthesis. Alternatively, solubilizing tags can be attached to the protein to promote soluble expression. Some commonly used solubilizng tags include glutathione s-transferase (GST), maltose binding protein (MBP), small ubiquitin-like modifier (SUMO), or N-utilizing substance A (NusA)[75]. Expression with these tags helps solubilize the fused protein target, but can reduce total protein yield. Furthermore, removal of the fused tag might be required. Chaperone-assisted protein folding is also utilized to help solubilize recombinant proteins; however, the target protein still must be

separated from the chaperone[75,77]. Unfortunately, these strategies often nullify the time and cost-effectiveness advantages inherent to bacterial over-expression systems.

As an alternative approach, investigators can disregard solubility issues, intentionally drive the protein into the inclusion body, and prioritize maximizing yields obtained for the protein target. Subsequently, protein in the inclusion body can be solubilized and refolded. Current *in vitro* strategies for protein refolding include direct dilution into refolding buffer[78,79], slow denaturant removal by dialysis[80,81], and chromatographic separation by size exclusion chromatography (SEC) or hydrophobic interaction chromatography (HIC)[82–84]. Protein refolding methodologies follow a similar approach – proteins trapped in amorphous aggregates or amyloid-like fibrils can be unfolded from kinetically trapped states to a higher energy transition state, thus allowing the protein another chance to fold properly (Figure 1-5)[85].

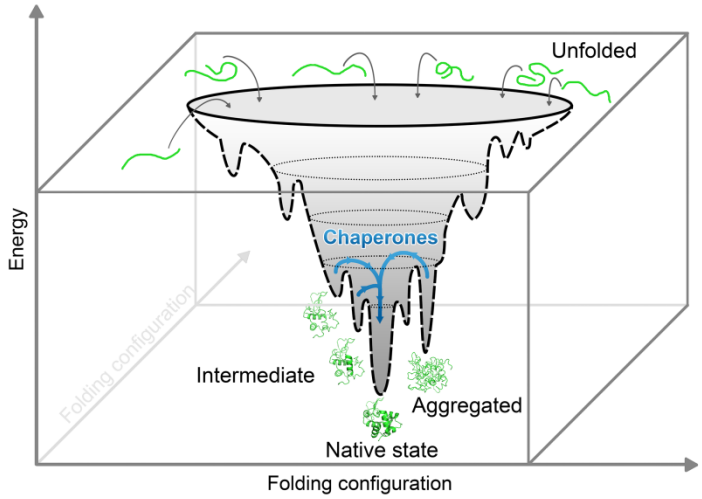


Figure 1-5. Protein folding in the folding funnel model. Unfolded proteins at the top of an energy landscape follow a free-energy surface in the shape of a rugged funnel, in which partially folded proteins can become kinetically trapped and form protein aggregates[70,86]. Prokaryotes and eukaryotes employ molecular chaperones to return misfolded and aggregated molecules to a higher free-energy state, allowing the protein to refold while isolated to reach the native state conformation[87–89]. Figure adapted from reference[86].

1.6 Refolding misfolded protein aggregates

Prokaryotes and eukaryotes have evolved similar proteins, collectively termed chaperones, to physically unfold protein aggregates[85,88,90,91]. Lin *et al*. demonstrated that in *E. coli*, forced unfolding of substrate protein is completed following binding of ATP to the *trans* GroEL-GroES ring[87]. Analogous mechanical *in vitro* approaches to unfolding protein aggregates are being actively explored. High hydrostatic pressures above 4 kbar separate oligomers and denature protein secondary structure[92,93]. Studies performed by John *et al.* demonstrate application of controlled hydrostatic pressures in the 1–2 kbar range to recover protein from inclusion body aggregates[94]. These

publications show that mechanical phenomena first applied to unravel natively folded protein can then be adapted, though controlled application, to unfold misfolded proteins into their native structures.

1.6.1 Shear stress mediates protein unfolding

Similarly, high shear stresses are also capable of denaturing protein secondary and tertiary structures. A shear force is applied to a liquid medium as the fluid moves along a stationary boundary with respect to the liquid (Figure 1-6).

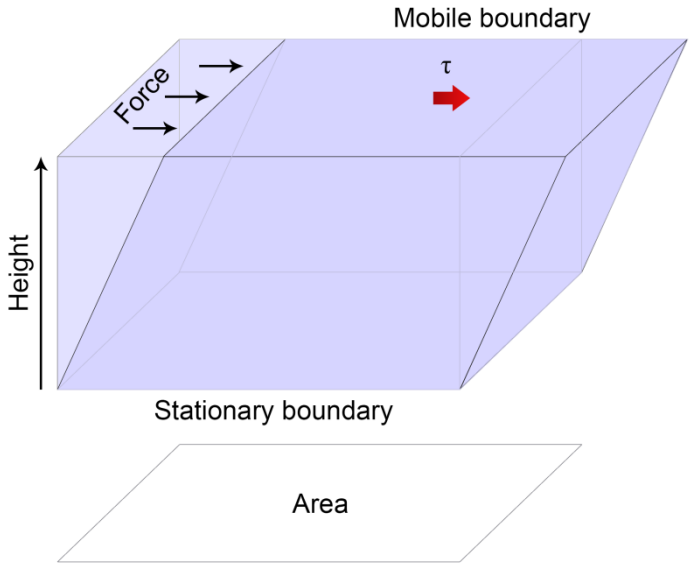


Figure 1-6. Simplified three dimensional rectangular section of fluid under shear stress (τ). Liquid in this system is moving to the right, relative to the stationary boundary. When the velocity gradient from the stationary boundary to the mobile boundary is linear, shear stress is defined as $\tau = \frac{F}{A}$, where A is the area to which force, F, is applied.

Shearing forces are often present as an unwanted byproduct of common biotechnology processes such as centrifugation, fractionation, and cell lysis[95–97]. Previous experiments quantified shear stress levels necessary to unfold and aggregate

insulin[98] and recombinant human growth hormone (rhGH)[99]. Experiments performed by Hill *et al.* demonstrated high shear stresses inducing amyloid fibril formation in β-lactoglobulin[100]. Similarily, controlled application of high hydrostatic pressures can both denature native protein and refold misfolded protein substrates, fine-tuning the amount and duration of shear stresses can also encourage unfolding and proper refolding to the native state without requiring a complex pressurized setup.

1.6.2 Controlled application of shear stress to liquid mediums

Researchers in the Colin Raston lab (Flinders University) developed a low-cost, benchtop device called a vortex fluidic device (VFD) to apply controlled levels of shear stress for the purpose of exfoliating graphite and hexagonal boron-nitride (Figure 1-7)[101,102]. Shear forces were shown to overcome substantial van der Waals interactions between the layers but did not damage the resulting monolayers – a common limitation associated with sonication and high pressure techniques. Controlled application of shear forces lead to reversible dissociation of non-covalent bonds as demonstrated by disassembly of hydrogen-bonded *p*-phosphonic acid calix[5]arenes[103]. Our lab then collaborated with Professor Raston and coworkers to report the first known application of applying shear stress to unfold protein aggregates (unpublished results).

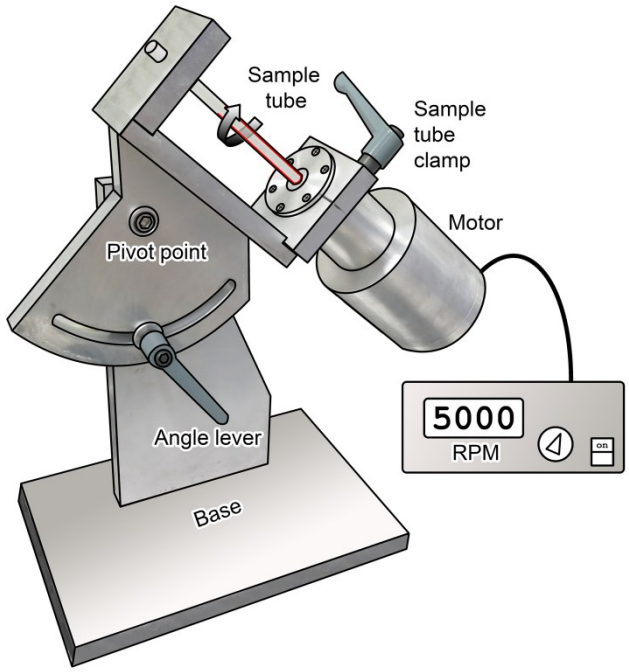


Figure 1-7. Schematic representation of a bench-top vortex fluid device (VFD). A 10 mm diameter glass tube, 16 cm long, and inclined at a 45° angle is held in place. When spun at high speeds, the liquid inside the sample tube is forced into thin microfluidic films ranging from 0.2 – 1.2 mm for a 1 mL volume (red outline). The thickness of the film is dependent on the rotational speed and angle of the tube[104].

1.7 Expanding applications for directed evolution

Conventional screening methods typically require long, time-consuming library generation and synthesis steps prior to the screening process. Mutagenesis is typically performed by techniques such as error-prone PCR, site-saturation mutagenesis, loop insertion or deletion, and homologous recombination[16]. However, nature already provides an optimized solution to generate extensive sequence diversity on receptor-binding scaffolds[29,42]. While vertebrate antibody fragments are successfully adapted to directed evolution strategies, CDR diversification is still inaccessible *in vitro*. Alternatively, the recent discovery of DGRs allows easy bench-top diversification of the generalized protein-binding ligand Mtd protein[56,105,106].

New techniques are continuously being developed to exploit protein-based ligands. Bispecific monoclonal antibodies represent a relatively new class of antibody conjugates that binds to two separate antigens[107]. Overstreet *et al*. recently explored expansion of the DGR to expand Brt-mediated library diversity[108]. Computational advances have predicted binding energy hot spots among protein-protein interfaces[109], and provided guidelines for selection stringencies to identify ligands with low off-rates[110].

Advances have also been made towards understanding protein aggregation kinetics in the crowded cytosolic environment of proteins, carbohydrates, nucleic acids and lipids[111]. Further investigation in refolding techniques for the remaining protein targets resistant to soluble expression include using mixed micelle and detergent solutions[112] or temperature-dependent hydrophobic polymers[113]. Cumulative advances in sequencing methods have led to a considerable decrease in the costs associated with sequencing entire genomes and spawned entirely new industries[114]. Similarly, reducing the cost of entry to protein engineering efforts could enable novel applications that have been prohibitively time-consuming or expensive.

REFERENCES

1. Dobson, C., Šali, A. & Karplus, M. Protein Folding: A Perspective From Theory and Experiment. *Angew. Chemie Int. Ed.* **37,** 868–893 (1998).

2. Nelson, D. L. & Cox, M. M. in *Lehninger Princ. Biochem.* **4th edition,** 1100 (2005).

3. Estell, D. A. A., Graycar, T. P. & Wells, J. A. Engineering an Enzyme by Site-directed Mutagenesis to Be Resistant to Chemical Oxidation. *J. Biol. Chem.* **260,** 6518–6521 (1985).

4. Vigneri, R., Squatrito, S. & Sciacca, L. Insulin and its analogs: actions via insulin and IGF receptors. *Acta Diabetol.* **47,** 271–278 (2010).

5. Hemminki, a *et al.* Specificity improvement of a recombinant anti-testosterone Fab fragment by CDRIII mutagenesis and phage display selection. *Protein Eng.* **11,** 311–319 (1998).

6. Carter, P. J. Introduction to current and future protein therapeutics: a protein engineering perspective. *Exp. Cell Res.* **317,** 1261–1269 (2011).

7. Walsh, G. Biopharmaceutical benchmarks 2010. *Nat. Biotechnol.* **28,** 917–924 (2010).

8. Clackson, T. & Wells, J. a. A hot spot of binding energy in a hormone-receptor interface. *Science.* **267,** 383–386 (1995).

9. Nuttall, S. D. & Walsh, R. B. Display scaffolds: protein engineering for novel therapeutics. *Curr Opin Pharmacol.* **8,** 609–615 (2008).

10. Brannigan, J. a & Wilkinson, A. J. Protein engineering 20 years on. *Nat. Rev. Mol. Cell Biol.* **3,** 964–970 (2002).

11. Reichert, J. M. Metrics for antibody therapeutics development. *MAbs.* **2,** 695–700 (2010).

12. Kelley, B. Industrialization of mAb production technology: the bioprocessing industry at a crossroads. *MAbs.* **1,** 443–452 (2009).

13. Wymer, N. *et al.* Directed evolution of a new catalytic site in 2-keto-3-deoxy-6-phosphogluconate aldolase from Escherichia coli. *Structure.* **9,** 1–9 (2001).

14. Fong, S., Machajewski, T. D., Mak, C. C. & Wong, C. H. Directed evolution of D-2-keto-3-deoxy-6-phosphogluconate aldolase to new variants for the efficient synthesis of D- and L-sugars. *Chem. Biol.* **7,** 873–883 (2000).

15. Windle, C. L., Müller, M., Nelson, A. & Berry, A. Engineering aldolases as biocatalysts. *Curr. Opin. Chem. Biol.* **19C,** 25–33 (2014).

16. Kazlauskas, R. J. & Bornscheuer, U. T. Finding better protein engineering strategies. *Nat. Chem. Biol.* **5,** 526–529 (2009).

17. Zhao, J., Kardashliev, T., Ruff, A. J., Bocola, M. & Schwaneberg, U. Lessons from diversity of directed evolution experiments by an analysis of 3000 mutations. *Biotechnol. Bioeng.* (2014). doi:10.1002/bit.25302

18. Bornscheuer, U. T. & Pohl, M. Improved biocatalysts by directed evolution and rational protein design. *Curr. Opin. Chem. Biol.* **5,** 137–143 (2001).

19. Tao, H. & Cornish, V. W. Milestones in directed enzyme evolution. *Curr. Opin. Chem. Biol.* **6,** 858–864 (2002).

20. Dror, A., Shemesh, E., Dayan, N. & Fishman, A. Protein engineering by random mutagenesis and structure-guided consensus of Geobacillus stearothermophilus Lipase T6 for enhanced stability in methanol. *Appl. Environ. Microbiol.* **80,** 1515–1527 (2014).

21. Lauchli, R. *et al.* High-throughput screening for terpene-synthase-cyclization activity and directed evolution of a terpene synthase. *Angew. Chemie - Int. Ed.* **52,** 5571–5574 (2013).

22. Tee, K. L. & Wong, T. S. Polishing the craft of genetic diversity creation in directed evolution. *Biotechnol. Adv.* **31,** 1707–1721 (2013).

23. Farinas, E. T., Bulter, T. & Arnold, F. H. Directed enzyme evolution. *Curr. Opin. Biotechnol.* **12,** 545–551 (2001).

24. Gascoigne, N. R. J. & Zal, T. Molecular interactions at the T cell-antigen-presenting cell interface. *Curr. Opin. Immunol.* **16,** 114–119 (2004).

25. Reichmann, D., Rahat, O., Cohen, M., Neuvirth, H. & Schreiber, G. The molecular architecture of protein – protein binding sites. doi:10.1016/j.sbi.2007.01.004

26. Reichmann, D., Rahat, O., Cohen, M., Neuvirth, H. & Schreiber, G. The molecular architecture of protein-protein binding sites. *Curr. Opin. Struct. Biol.* **17,** 67–76 (2007).

27. Lo Conte, L., Chothia, C. & Janin, J. The atomic structure of protein-protein recognition sites. *J. Mol. Biol.* **285,** 2177–2198 (1999).

28. Norel, R., Lin, S. L., Wolfson, H. J. & Nussinov, R. Shape complementarity at protein-protein interfaces. *Biopolymers.* **34,** 933–940 (1994).

29. Mian, I. S., Bradwell, A. R. & Olson, A. J. Structure, function and properties of antibody binding sites. *J. Mol. Biol.* **217,** 133–151 (1991).

30. Deng, L. *et al.* A structural basis for antigen recognition by the T cell-like lymphocytes of sea lamprey. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 13408–13413 (2010).

31. Chothia, C. & Lesk, A. M. Canonical structures for the hypervariable regions of immunoglobulins. *J. Mol. Biol.* **196,** 901–17 (1987).

32. Davis, M. M. & Bjorkman, P. J. T-cell antigen receptor genes and T-cell recognition. *Nature.* **334,** 395–402 (1988).

33. Schatz, D. G. & Ji, Y. Recombination centres and the orchestration of V(D)J recombination. *Nat. Rev. Immunol.* **11,** 251–263 (2011).

34. Teng, G. & Papavasiliou, F. N. Immunoglobulin somatic hypermutation. *Annu. Rev. Genet.* **41,** 107–120 (2007).

35. Pancer, Z. *et al.* Somatic diversification of variable lymphocyte receptors in the agnathan sea lamprey. *Nature* **430,** 174–180 (2004).

36. Pancer, Z. *et al.* Variable lymphocyte receptors in hagfish. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 9224–9229 (2005).

37. Binz, H. K. *et al.* High-affinity binders selected from designed ankyrin repeat protein libraries. *Nat. Biotechnol.* **22,** 575–582 (2004).

38. Binz, H. K. & Pluckthun, A. Engineered proteins as specific binding reagents. *Curr. Opin. Biotechnol.* **16,** 459–469 (2005).

39. Kohl, A. *et al.* Allosteric inhibition of aminoglycoside phosphotransferase by a designed ankyrin repeat protein. *Structure.* **13,** 1131–1141 (2005).

40. Wikman, M. *et al.* Selection and characterization of HER2/neu-binding affibody ligands. *Protein Eng. Des. Sel.* **17,** 455–62 (2004).

41. Harris, L. J., Larson, S. B., Hasel, K. W. & McPherson, A. Refined structure of an intact IgG2a monoclonal antibody. *Biochemistry.* **36,** 1581–1597 (1997).

42. McMahon, S. A. *et al.* The C-type lectin fold as an evolutionary solution for massive sequence variation. *Nat. Struct. Mol. Biol.* **12,** 886–892 (2005).

43. Mosavi, L. K., Minor, D. L. & Peng, Z.-Y. Consensus-derived structural determinants of the ankyrin repeat motif. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 16029–16034 (2002).

44. Lendel, C., Dogan, J. & Härd, T. Structural Basis for Molecular Recognition in an Affibody:Affibody Complex. *J. Mol. Biol.* **359,** 1293–1304 (2006).

45. Nelson, A. L. Antibody fragments: hope and hype. *MAbs.* **2,** 77–83 (2010).

46. Richie, D. L., Ghannoum, M. A., Isham, N., Thompson, K. V. & Ryder, N. S. Nonspecific Effect of Mycograb on Amphotericin B MIC. *Antimicrob. Agents Chemother.* **56,** 3963–3964 (2012).

47. Söderlind, E. *et al.* Recombining germline-derived CDR sequences for creating diverse single-framework antibody libraries. *Nat. Biotechnol.* **18,** 852–856 (2000).

48. Holt, L. J., Herring, C., Jespers, L. S., Woolven, B. P. & Tomlinson, I. M. Domain antibodies: proteins for therapy. *Trends Biotechnol.* **21,** 484–490 (2003).

49. Andersen, D. C. & Krummen, L. Recombinant protein expression for therapeutic applications. *Curr. Opin. Biotechnol.* **13,** 117–123 (2002).

50. Guo, H. *et al.* Diversity-generating retroelement homing regenerates target sequences for repeated rounds of codon rewriting and protein diversification. *Mol. Cell.* **31,** 813–823 (2008).

51. Neylon, C. Chemical and biochemical strategies for the randomization of protein encoding DNA sequences: library construction methods for directed evolution. *Nucleic Acids Res.* **32,** 1448–1459 (2004).

52. Liu, M. *et al.* Reverse transcriptase-mediated tropism switching in Bordetella bacteriophage. *Science.* **295,** 2091–2094 (2002).

53. Miller, J. F. J. L. F., Le Coq, J., Hodes, A., Barbalat, R. & Ghosh, P. Selective ligand recognition by a diversity-generating retroelement variable protein. *PLoS Biol.* **6,** e131 (2008).

54. Bock, A. & Gross, R. The BvgAS two-component system of Bordetella spp.: a versatile modulator of virulence gene expression. *Int. J. Med. Microbiol.* **291,** 119–130 (2001).

55. Akerley, B. J., Cotter, P. A. & Miller, J. F. Ectopic expression of the flagellar regulon alters development of the Bordetella-host interaction. *Cell* **80,** 611–620 (1995).

56. Liu, M. *et al.* Tropism switching in Bordetella bacteriophage defines a family of diversity-generating retroelements. *Nature* **431,** 476–481 (2004).

57. Hester, G., Kaku, H., Goldstein, I. J. & Wright, C. S. Structure of mannose-specific snowdrop (Galanthus nivalis) lectin is representative of a new plant lectin family. *Nat. Struct. Biol.* **2,** 472–479 (1995).

58. Sauerborn, M. K., Wright, L. M., Reynolds, C. D., Grossmann, J. G. & Rizkallah, P. J. Insights into carbohydrate recognition by Narcissus pseudonarcissus lectin: the crystal structure at 2 A resolution in complex with alpha1-3 mannobiose. *J. Mol. Biol.* **290,** 185–199 (1999).

59. Luo, Y. *et al.* Crystal structure of enteropathogenic Escherichia coli intimin-receptor complex. *Nature.* **405,** 1073–1077 (2000).

60. Weis, W. I., Kahn, R., Fourme, R., Drickamer, K. & Hendrickson, W. A. Structure of the calcium-dependent lectin domain from a rat mannose-binding protein determined by MAD phasing. *Science.* **254,** 1608–1615 (1991).

61. Dai, W. *et al.* Three-dimensional structure of tropism-switching Bordetella bacteriophage. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 4347–4352 (2010).

62. Smith, G. P. & Petrenko, V. a. Phage Display. *Chem Rev* **97,** 391–410 (1997).

63. Gram, H. *et al.* In vitro selection and affinity maturation of antibodies from a naive combinatorial immunoglobulin library. *Proc. Natl. Acad. Sci. U. S. A.* **89,** 3576–3580 (1992).

64. Hoogenboom, H. R. *et al.* Multi-subunit proteins on the surface of filamentous phage: methodologies for displaying antibody (Fab) heavy and light chains. *Nucleic Acids Res.* **19,** 4133–4137 (1991).

65. Garrard, L. J., Yang, M., O'Connell, M. P., Kelley, R. F. & Henner, D. J. Fab assembly and enrichment in a monovalent phage display system. *Biotechnology. (N. Y).* **9,** 1373–1377 (1991).

66. McCafferty, J., Griffiths, A. D., Winter, G. & Chiswell, D. J. Phage antibodies: filamentous phage displaying antibody variable domains. *Nature.* **348,** 552–554 (1990).

67. Chowdhury, P. S. & Pastan, I. Improving antibody affinity by mimicking somatic hypermutation in vitro. *Nat. Biotechnol.* **17,** 568–572 (1999).

68. Braren, I. *et al.* Generation of human monoclonal allergen-specific IgE and IgG antibodies from synthetic antibody libraries. *Clin. Chem.* **53,** 837–844 (2007).

69. Yuan, T. Z., Overstreet, C. M., Moody, I. S. & Weiss, G. A. Protein Engineering with Biosynthesized Libraries from Bordetella bronchiseptica Bacteriophage. *PLoS One.* **8,** e55617 (2013).

70. Radford, S. E. Protein folding: progress made and promises ahead. *Trends Biochem. Sci.* **25,** 611–618 (2000).

71. Jemth, P. *et al.* Demonstration of a low-energy on-pathway intermediate in a fast-folding protein by kinetics, protein engineering, and simulation. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 6450–6455 (2004).

72. Wang, L., Schubert, D., Sawaya, M. R., Eisenberg, D. & Riek, R. Multidimensional structure-activity relationship of a protein in its aggregated states. *Angew. Chem. Int. Ed. Engl.* **49,** 3904–3908 (2010).

73. Wang, L., Maji, S. K., Sawaya, M. R., Eisenberg, D. & Riek, R. Bacterial inclusion bodies contain amyloid-like structure. *PLoS Biol.* **6,** e195 (2008).

74. Hannig, G. & Makrides, S. C. Strategies for optimizing heterologous protein expression in Escherichia coli. *Trends Biotechnol.* **16,** 54–60 (1998).

75. Francis, D. M. & Page, R. Strategies to optimize protein expression in E. coli. *Curr. Protoc. Protein Sci.* **Chapter 5,** Unit 5.24.1–29 (2010).

76. Berrow, N. S. *et al.* Recombinant protein expression and solubility screening in *Escherichia coli*: a comparative study. *Acta Crystallogr. D. Biol. Crystallogr.* **62,** 1218–1226 (2006).

77. Haacke, A., Fendrich, G., Ramage, P. & Geiser, M. Chaperone over-expression in Escherichia coli: apparent increased yields of soluble recombinant protein kinases are due mainly to soluble aggregates. *Protein Expr. Purif.* **64,** 185–193 (2009).

78. Katoh, S. & Katoh, Y. Continuous refolding of lysozyme with fed-batch addition of denatured protein solution. *Process Biochem.* **35,** 1119–1124 (2000).

79. Fischer, B., Perry, B., Sumner, I. & Goodenough, P. A novel sequential procedure to enhance the renaturation of recombinant protein from Escherichia coli inclusion bodies. *Protein Eng.* **5,** 593–596 (1992).

80. Maeda, Y., Koga, H., Yamada, H., Ueda, T. & Imoto, T. Effective renaturation of reduced lysozyme by gentle removal of urea. *Protein Eng.* **8,** 201–205 (1995).

81. West, S. M., Chaudhuri, J. B. & Howell, J. A. Improved protein refolding using hollow-fibre membrane dialysis. *Biotechnol. Bioeng.* **57,** 590–599 (1998).

82. De Bernardez Clark, E., Schwarz, E. & Rudolph, R. Inhibition of aggregation side reactions during in vitro protein folding. *Methods Enzymol.* **309,** 217–236 (1999).

83. Geng, X. & Chang, X. High-performance hydrophobic interaction chromatography as a tool for protein refolding. in *J. Chromatogr.* **599,** 185–194 (1992).

84. Vallejo, L. F. & Rinas, U. Strategies for the recovery of active proteins through refolding of bacterial inclusion body proteins. *Microb. Cell Fact.* **3,** 11 (2004).

85. Hartl, F. U. & Hayer-Hartl, M. Converging concepts of protein folding in vitro and in vivo. *Nat. Struct. Mol. Biol.* **16,** 574–581 (2009).

86. Schultz, C. P. Illuminating folding intermediates. *Nat. Struct. Biol.* **7,** 7–10 (2000).

87. Lin, Z., Madan, D. & Rye, H. S. GroEL stimulates protein folding through forced unfolding. *Nat. Struct. Mol. Biol.* **15,** 303–311 (2008).

88. Hartl, F. U., Bracher, A. & Hayer-Hartl, M. Molecular chaperones in protein folding and proteostasis. *Nature.* **475,** 324–332 (2011).

89. Chakraborty, K. *et al.* Chaperonin-catalyzed rescue of kinetically trapped states in protein folding. *Cell.* **142,** 112–122 (2010).

90. Hemmingsen, S. *et al.* Homologous plant and bacterial proteins chaperone oligomeric protein assembly. *Nature.* **333,** 330–334 (1988).

91. Hartl, F. Molecular chaperones in cellular protein folding. *Nature.* **381,** 571–579 (1996).

92. Gorovits, B. M. & Horowitz, P. M. High hydrostatic pressure can reverse aggregation of protein folding intermediates and facilitate acquisition of native structure. *Biochemistry.* **37,** 6132–6135 (1998).

93. Ruan, K. & Weber, G. Dissociation of yeast hexokinase by hydrostatic pressure. *Biochemistry.* **27,** 3295–3301 (1988).

94. St John, R. J., Carpenter, J. F. & Randolph, T. W. High pressure fosters protein refolding from aggregates at high concentrations. *Proc. Natl. Acad. Sci. U. S. A.* **96,** 13029–13033 (1999).

95. Yim, S. S. & Shamlou, P. A. The engineering effects of fluids flow on freely suspended biological macro-materials and macromolecules. *Adv. Biochem. Eng. Biotechnol.* **67,** 83–122 (2000).

96. Lin, J. J., Meyer, J. D., Carpenter, J. F. & Manning, M. C. Stability of human serum albumin during bioprocessing: Denaturation and aggregation during processing of albumin paste. *Pharm. Res.* **17,** 391–396 (2000).

97. Bekard, I. B., Asimakis, P., Bertolini, J. & Dunstan, D. E. The effects of shear flow on protein structure and function. *Biopolymers.* **95,** 733–745 (2011).

98. Bekard, I. B. & Dunstan, D. E. Shear-induced deformation of bovine insulin in Couette flow. *J. Phys. Chem. B.* **113,** 8453–8457 (2009).

99. Maa, Y. F. & Hsu, C. C. Effect of high shear on proteins. *Biotechnol. Bioeng.* **51,** 458–465 (1996).

100. Hill, E. K., Krebs, B., Goodall, D. G., Howlett, G. J. & Dunstan, D. E. Shear flow induces amyloid fibril formation. *Biomacromolecules.* **7,** 10–13 (2006).

101. Chen, X., Dobson, J. F. & Raston, C. L. Vortex fluidic exfoliation of graphite and boron nitride. *Chem. Commun.* **48,** 3703–3705 (2012).

102. Wahid, M. H., Eroglu, E., Chen, X., Smith, S. S. M. & Raston, C. C. L. Functional multi-layer graphene–algae hybrid material formed using vortex fluidics. *Green Chem.* **15,** 650–655 (2013).

103. Martin, A. D., Boulos, R. A., Hubble, L. J., Hartlieb, K. J. & Raston, C. L. Multifunctional water-soluble molecular capsules based on p-phosphonic acid calix[5]arene. *Chem. Commun.* **47,** 7353–7355 (2011).

104. Yasmin, L., Chen, X., Stubbs, K. a. & Raston, C. L. Optimising a vortex fluidic device for controlling chemical reactivity and selectivity. *Sci. Rep.* **3,** 2282 (2013).

105. Guo, H. *et al.* Target site recognition by a diversity-generating retroelement. *PLoS Genet.* **7,** e1002414 (2011).

106. Medhekar, B. & Miller, J. F. Diversity-generating retroelements. *Curr. Opin. Microbiol.* **10,** 388–395 (2007).

107. Baeuerle, P. a & Reinhardt, C. Bispecific T-cell engaging antibodies for cancer therapy. *Cancer Res.* **69,** 4941–4944 (2009).

108. Overstreet, C. M., Levin, A. M., Kong, C., Coroneus, J. G. & Weiss, G. A. Harnessing a Self-Made Phage Library. 1–22

109. Kortemme, T. & Baker, D. A simple physical model for binding energy hot spots in protein-protein complexes. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 14116–14121 (2002).

110. Zahnd, C., Sarkar, C. A., Plu, A., Pluckthun, A. & Plückthun, A. Computational analysis of off-rate selection experiments to optimize affinity maturation by directed evolution. *Protein Eng. Des. Sel.* **23,** 1–10 (2010).

111. White, D. A., Buell, A. K., Knowles, T. P. J., Welland, M. E. & Dobson, C. M. Protein aggregation in crowded environments. *J. Am. Chem. Soc.* **132,** 5170–5175 (2010).

112. Zardeneta, G. & Horowitz, P. M. Protein refolding at high concentrations using detergent/phospholipid mixtures. *Anal. Biochem.* **218,** 392–398 (1994).

113. Kuboi, R., Morita, S., Ota, H. & Umakoshi, H. Protein refolding using stimuli-responsive polymer-modified aqueous two- phase systems. *J. Chromatogr. B Biomed. Sci. Appl.* **743,** 215–223 (2000).

114. Schuster, S. C. Next-generation sequencing transforms today's biology. *Nat. Methods* **5,** 16–18 (2008).

CHAPTER 2

Protein engineering with biosynthesized libraries from *Bordetella bronchiseptica* bacteriophage

Abstract

Phage display offers a powerful approach to engineer protein affinity. A naturally occurring analog to phage display, the *Bordetella bronchiseptica* bacteriophage (BP) employs a highly variable protein termed the major tropism determinant (Mtd) to recognize its dynamic host. Propagation of BP provides a self-made phage library (SMPL) with vast numbers of phage particles, each displaying a single Mtd variant. We report applying the diversity of the BP-SMPL to access a tyrosine-rich library of Mtd variants. Expression of the SMPL-engineered Mtd variant as a GST-bound fusion protein demonstrated specific binding to the target T4 lysozyme with dissociation constants in the sub-micromolar range. The results guide future experiments with SMPLs applied to protein engineering.

Introduction

In vitro molecular evolution is extensively used for the identification and optimization of binding by receptors and biopharmaceuticals[1–4]. Such experiments take cues from the immune system, and offer rapid evolution of high affinity binding proteins. *Bordetella bronchiseptica* bacteriophage (BP) have evolved a diversity generating retroelement, which synthesizes self-made phage libraries (SMPLs) by introducing DNA mutations into the gene encoding the major tropism determinant (Mtd) protein on the tail fibers of each BP[5,6]. The BP-SMPLs demonstrate various attributes found in effective molecular display systems - vast diversity, flexible binding to a range of targets, and encapsulated sequence information.

The Mtd protein also determines viral specificity for its host, *B. bronchiseptica*, by binding to outer membrane proteins expressed on the bacterial surface[7]. The bacterial host alters its surface proteins during transitions between virulent (Bvg^+) and avirulent (Bvg^-) phases of its life cycle. *Bordetella* bacteria exist primarily in this bimodal phase system. Activation of the *Bordetella* virulence control locus *BvgAS* enables expression of virulence factors in the Bvg^+ phase. Separate genes are activated to enable phage motility in the environmental, avirulent Bvg^- phase[8–10]. To maintain infectivity, the BP diversity generating retroelement actively mutates the DNA sequence encoding the C-terminus of the Mtd (Figure 2-1). During phage propagation, the phage produces a BP-SMPL consisting of a vast library of Mtd variants. A subset of Mtd variants allows the BP to switch tropism, and bind to the new phase of the host. The BP-derived SMPL allows the phage to maintain its infectivity for a dynamically changing host.

Biosynthesis of the BP-SMPL relies upon a phage-encoded, error-prone reverse transcriptase. Sequence information from a non-coding template region (*tr*) of the phage genome is transferred to the variable region (*vr*) at the 3' terminus of the *mtd* gene, which encodes the C-terminus of the Mtd. Before this transfer, *Bordetella* reverse transcriptase mutates the *tr* mRNA, substituting adenines with any of the four DNA bases[5,7,11]. These adenine-dependent mutations correspond to twelve codons located in the coding *vr* region. In its native form, the BP-SMPL is theoretically capable of encoding up to 9.2 x 10^{12} unique Mtd variants (Figure 2-1B).

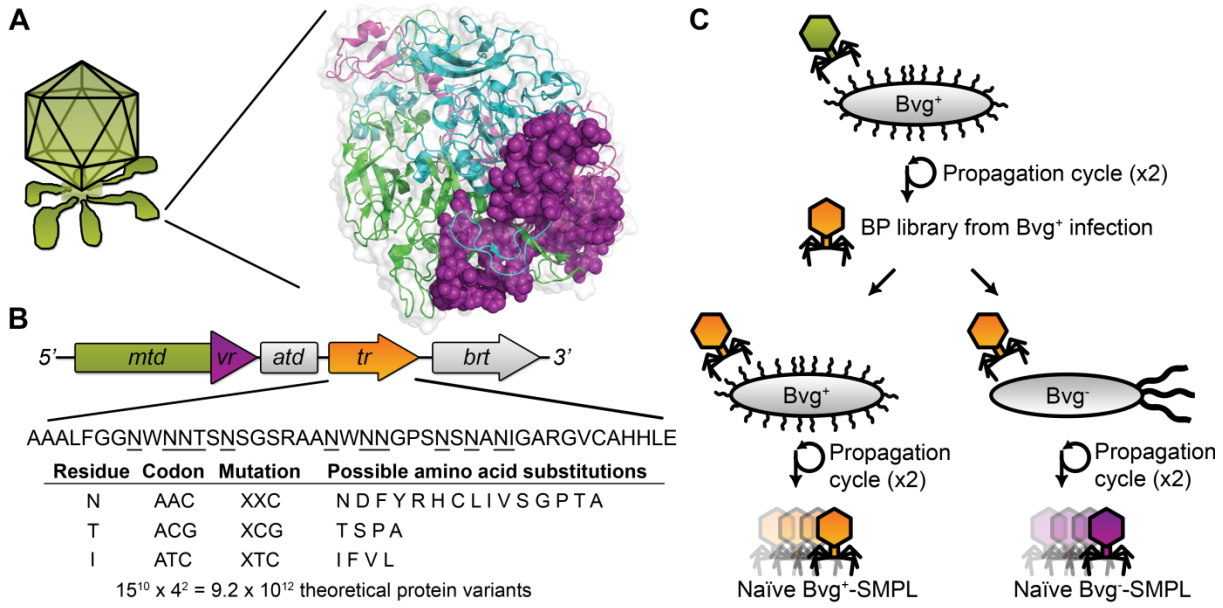


Figure 2-1. Mtd protein structure and SMPL diversity generation. (A) *Bordetella* phage expresses six distal tail fibers, with the Mtd protein located at the end of each tail fiber. The Mtd structure consists of a homotrimer found on the tail fibers of the BP (PDB: 1YU4)[12]. During SMPL formation, the variable region (VR) of the trimer (purple spheres) is diversified, and is responsible for determining binding specificity to target proteins on the surface of *Bordetella bronchiseptica*[12,13]. (B) This schematic gene map of the diversity generating retroelement identifies mutable codons within the *tr*. The targeted codons include adenines outside the wobble position, and encode the underlined amino acids to generate 9.2 x 10^{12} variants of the Mtd protein. (C) Generation of the BP-SMPLs through phage production by infecting Bvg^+ *Bordetella bronchiseptica*, followed by propagation in either Bvg^+ or Bvg^- phase *Bordetella* strains.

The BP-SMPL offers tremendous library diversity in a more expedient format than conventional molecular display techniques due to the virus self-synthesizing a new protein library upon propagation in the bacterial host. The bias inherent to propagation could be either avoided or exploited, if characterized for different hosts. Here, we define the properties of SMPLs propagated in two hosts (Figure 2-1), and conduct selections targeting T4 lysozyme (subsequently referred to as lysozyme) with libraries produced by the BP (Figure 2-2). Protein binding assays with an expressed and purified variant, the lysozyme-binding Mtd (termed L-Mtd) from the BP-SMPL demonstrate the effectiveness of the BP-SMPL system for the identification of specific, high affinity binding partners (Figure 2-5).

2.1 Results and Discussion

The synthesis of a BP-SMPL from prophage integrated into the host Bvg^+ cells requires one or more cycles of phage infection, growth, and isolation. Each cycle can accumulate diverse Mtd sequences for potential tropism switching. Increasing the number of phage in the naïve library provides a more diverse starting population, as the SMPL relies on the host for library biosynthesis. BP produced from the Bvg^+ wild-type prophage were first propagated for two cycles in Bvg^+ phase bacteria to increase the viral titers (Figure 2-1C).

Following the two cycles of infection and growth in Bvg^+ bacteria to boost titers, two different SMPLs were generated by infecting either Bvg^+ or Bvg^- host cells. Each library was then amplified for an additional two cycles to accumulate diversity before selections. Two unique SMPLs resulted from propagation in either the Bvg^+ or Bvg^-

bacteria, and were designated as either a Bvg$^+$-SMPL or a Bvg$^-$-SMPL, respectively. During the four initial cycles of growth in phase-locked strains of *B. bronchiseptica*, the two SMPLs were not subjected to *in vitro* selections, and were thus considered naïve libraries.

2.1.1 Characterization of naïve *Bordetella* phage libraries

To characterize the two initial libraries, the *mtd* genes of naïve BP-SMPLs were PCR-amplified and sequenced. The majority of Mtd variants identified from the naïve Bvg$^-$-SMPL possessed unique *vr* sequences. In contrast, all clones sequenced from the naïve Bvg$^+$-SMPL were wild-type (Table 2-1). The lack of diversity in the Bvg$^+$-SMPL was expected, as typically only 1 in 10^6 of the propagated phage can acquire mutations enabling the switch in tropism from Bvg$^+$ to Bvg$^-$[14]. Amongst the 47 clones from the naïve Bvg$^-$-SMPL, 36 unique variations of the Mtd were observed (Table 2-1).

		***mtd vr* Position**												
Variant	**Library**	344	346	347	348	350	357	359	360	364	366	368	369	# of clones
Wild-type	-	A	N	G	T	L	L	Y	S	F	F	F	F	
1	Bvg$^+$-SMPL	A	N	G	T	L	L	Y	S	F	F	F	F	17
18	Bvg$^-$-SMPL	S	S	N	T	T	Y	R	N	N	Y	Y	I	1
19	Bvg$^-$-SMPL	S	P	N	T	N	Y	L	S	N	Y	Y	I	1
20	Bvg$^-$-SMPL	S	S	N	T	I	Y	T	D	G	N	Y	F	1
21	Bvg$^-$-SMPL	S	S	N	T	S	L	Y	S	Y	F	Y	L	1
22	Bvg$^-$-SMPL	S	S	N	T	N	Y	Y	S	Y	N	Y	I	1
23	Bvg$^-$-SMPL	A	S	N	T	N	Y	N	D	D	P	Y	I	1
24	Bvg$^-$-SMPL	A	S	N	T	A	Y	Y	N	F	F	F	F	1
25	Bvg$^-$-SMPL	S	S	R	T	N	S	Y	A	N	Y	G	I	1
26	Bvg$^-$-SMPL	S	T	N	T	N	Y	D	Y	A	A	Y	I	1
27	Bvg$^-$-SMPL	S	S	R	T	N	Y	N	N	S	G	G	V	1
28	Bvg$^-$-SMPL	S	S	N	T	S	Y	N	N	F	F	F	F	1
29	Bvg$^-$-SMPL	A	S	N	T	Y	Y	Y	S	F	F	F	F	1
30	Bvg$^-$-SMPL	A	S	N	T	Y	Y	Y	T	F	F	F	F	1
31	Bvg$^-$-SMPL	S	S	N	T	H	Y	N	I	L	Y	F	L	1
32	Bvg$^-$-SMPL	S	A	Y	T	Y	Y	A	S	S	A	N	V	1
33	Bvg$^-$-SMPL	Y	S	R	T	N	Y	N	S	N	S	G	I	1
34	Bvg$^-$-SMPL	S	S	N	T	H	Y	N	H	N	A	Y	I	1
35	Bvg$^-$-SMPL	Y	N	R	A	Y	Y	N	N	N	Y	G	I	1
36	Bvg$^-$-SMPL	S	A	Y	T	N	Y	N	S	S	A	N	I	1
37	Bvg$^-$-SMPL	S	S	R	T	Y	Y	N	Y	N	Y	Y	I	1
38	Bvg$^-$-SMPL	F	N	N	T	N	Y	N	N	Y	Y	Y	I	1
39	Bvg$^-$-SMPL	S	S	N	T	Y	Y	G	S	F	F	F	F	3
42	Bvg$^-$-SMPL	S	L	N	T	P	Y	Y	Y	Y	T	Y	I	1
44	Bvg$^-$-SMPL	A	S	N	T	N	Y	N	N	F	F	F	F	7
53	Bvg$^-$-SMPL	A	S	N	T	N	Y	H	Y	F	F	F	F	1
54	Bvg$^-$-SMPL	A	S	N	T	L	L	Y	T	F	F	F	F	1
55	Bvg$^-$-SMPL	A	S	N	T	L	L	N	S	F	F	F	F	1
56	Bvg$^-$-SMPL	A	S	N	T	N	Y	Y	Y	T	S	F	I	1
57	Bvg$^-$-SMPL	A	Y	N	T	N	Y	S	Y	N	Y	F	I	1
58	Bvg$^-$-SMPL	A	N	N	A	N	Y	S	Y	F	A	Y	I	1
59	Bvg$^-$-SMPL	S	S	N	T	Y	F	S	Y	A	Y	Y	I	1
60	Bvg$^-$-SMPL	S	S	N	T	Y	Y	S	Y	N	L	Y	I	1
61	Bvg$^-$-SMPL	S	S	N	T	N	Y	N	Y	A	Y	Y	I	1
62	Bvg$^-$-SMPL	S	S	N	T	Y	Y	S	I	F	Y	Y	I	1
63	Bvg$^-$-SMPL	S	S	N	T	Y	Y	Y	Y	N	Y	Y	I	1
64	Bvg$^-$-SMPL	A	L	N	T	N	N	Y	Y	S	Y	Y	I	1

Table 2-1. Mtd sequences of naïve Bvg$^+$-SMPL and naïve Bvg$^-$-SMPL variants. The yellow highlighting indicates deviation from the wild-type prophage sequence.

Since SMPLs are biosynthesized, the library volume, phage titers, and percent variation determine the diversity of the library. The diversity of conventional phage display libraries, by comparison, is determined by the transformation efficiency of library DNA into *E. coli* bacteria. The naïve Bvg$^-$-SMPL produced 3 ml of phage with a titer of

1.2 x 10^8 plaque forming units (PFU) per ml to yield an estimated diversity of 3.6 x 10^8 Mtd variants (Table 2-2). The titers and consequent library diversity could be expanded considerably using large scale growth conditions.

Library	Theoretical diversity	Titers (PFU/ml)	Practical diversity
Bordetella phage	9.2 x 10^{12}	1.2 x 10^8	3.6 x 10^8
M13 phage protein library	Varies	10^8 - 10^{10}	Up to 10^{10}
PCR-driven library construction (Antibody sourced)	Up to 10^9	-	Up to 10^9
Kunkel mutagenesis	Up to 10^{10}	-	Up to 10^{10}

Table 2-2. Practical and theoretical diversities from BP-SMPL production and comparable M13 libraries[2].

2.1.2 Selections with the BP-SMPLs

To demonstrate the effectiveness of BP-SMPL for the identification of high affinity binders, selections targeted T4 lysozyme. Our lab uses this protein in single molecule studies[15,16], and its binding partners and inhibitors could provide useful tools for biophysical studies. The Bvg^--SMPL was biopanned against biotinylated lysozyme in solution before capture with streptavidin-coated magnetic beads. Following short target incubation and bead capture times, competing non-biotinylated lysozyme (100-fold molar excess) was added to the solution to remove BP variants with fast off-rates. The selections for slow off-rates targeted a single cysteine variant of lysozyme biotinylated through conjugation to biotin-*N*-maleimide. PCR of the selected BP variants followed by sequencing of the *mtd* gene identified five unique Mtd variants. The selectant termed L-Mtd (for lysozyme binding Mtd) comprised 77% of the variants from the off-rate based

selections. The results demonstrate convergence on a selectant that is energetically favorable for binding in solution (Figure 2-2A).

2.1.3 Accessing a tyrosine-rich library with Mtd variants

Antibody complementarity-determining regions (CDRs) include flexible loops for the recognition of antigens; this structural flexibility can readily accommodate diverse binding partners within an otherwise rigid immunoglobulin domain recognition[17]. The Mtd features a VR segment in which flexible loops compose roughly half the mutable residues[18]. While not structurally similar to antibodies, the Mtd also provides a versatile system for adaptive molecular recognition with a balance of structural rigidity, sequence diversity, and sufficient malleability, as demonstrated by the ability of the BP-SMPL to generate ligands to novel protein targets not normally encountered by the *Bordetella* phage. Such features are attractive for the identification of binding partners to new targets.

The amino acid tyrosine comprises up to 25% of the complementarity determining regions of functional antibodies[17]. The useful role of tyrosine as a relatively inflexible, hydrophobic, yet potentially hydrogen bonding sidechain has also been demonstrated in the M13 phage display of antibody Fab domains extensively substituted with a "binary code" of exclusively tyrosine and serine residues[19]. Tyrosine residues also enhance the effectiveness of other phage-displayed, Fab libraries[20]. The enrichment for tyrosine substitutions also occurs in the Mtd variants isolated through selections for their slow off-rates for binding to lysozyme.

For example, the L-Mtd has a high percentage of tyrosine residues clustered in the center of the solvent-exposed Mtd VR. L-Mtd and other selectants from the Bvg$^-$-SMPL leveraged a high percentage of tyrosine residues at position 359, 360, and 364 (Figure 2-2C). The L-Mtd variant retains the wild-type VR sequence from positions 344 to 350, and a hydrophilic amino acid substitution profile characterizes positions 364, 366, and 368 (Figure 2-2A). For further analysis of the binding properties of selectants from the Bvg$^-$-SMPL, a number of Mtd variants, including L-Mtd, were overexpressed and purified without incorporation into phage particles.

A

Clone	Library	Target	Mtd VR sequence 344	346	347	348	350	357	359	360	364	366	368	369
wild-type	-	-	A	N	G	T	L	L	Y	S	F	F	F	F
1	Bvg-	T4 Lys	**S**	N	**N**	T	**N**	**Y**	**G**	**H**	**N**	**Y**	**Y**	**I**
2	Bvg-	T4 Lys	A	N	G	T	L	**N**	**R**	**N**	**N**	**N**	**S**	**I**
3	Bvg-	T4 Lys	**S**	**S**	**N**	T	**Y**	**Y**	**N**	**Y**	**T**	**H**	**Y**	**I**
4	Bvg-	T4 Lys	**Y**	N	**R**	T	L	**Y**	Y	S	**Y**	**Y**	**G**	**I**
* 5-17	Bvg-	T4 Lys	A	N	G	T	L	**Y**	Y	**Y**	**Y**	**N**	**R**	**I**

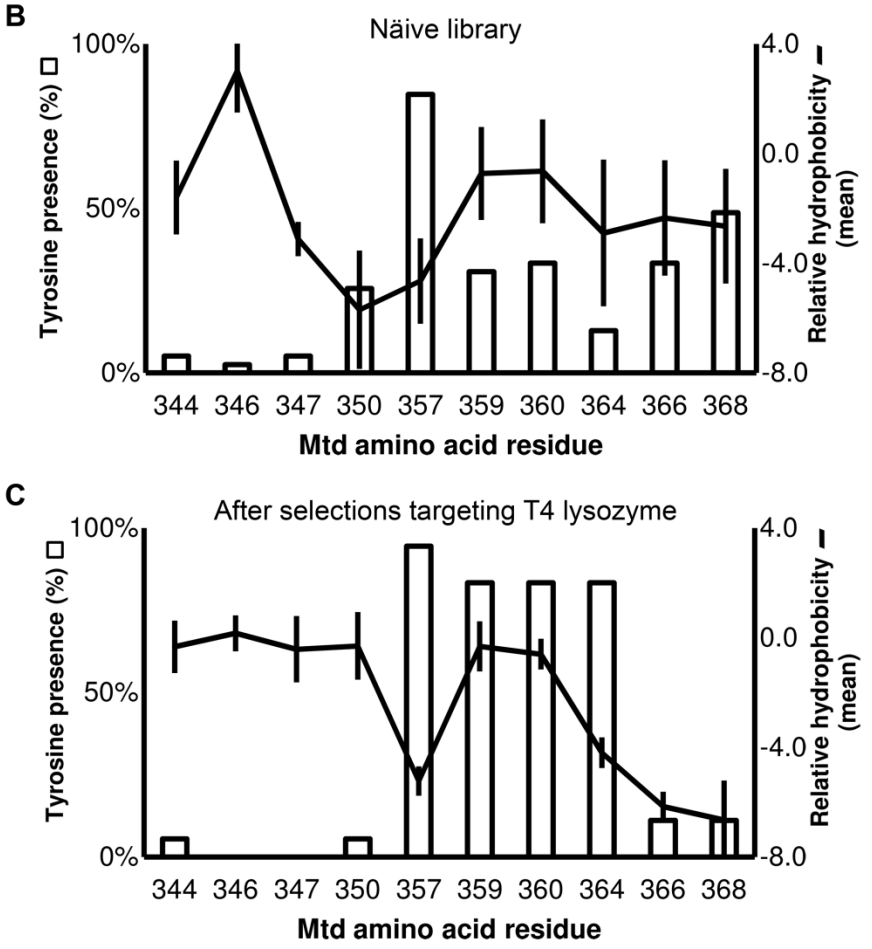


Figure 2-2. Mtd sequences and selections. (A) The mutable residues of clones recombinantly expressed as GST fusion proteins. The lysozyme-binding L-Mtd variant is indicated by the asterisk. Shading identifies altered residues based on the wild-type Mtd sequence. Analysis of sequences from the (B) naïve Bvg$^-$-SMPL library and sequences from (C) selections targeting lysozyme. The bar graph indicates the occurrence of tyrosine residues amongst the Bvg$^-$-SMPL sequences. The line graph indicates mean hydrophobicity scores for the Bvg$^-$-SMPL sequences relative to wild-type residues. Error bars represent standard deviation (n=36).

2.2 Characterization of Mtd-GST fusion proteins

The wild-type Mtd and five Mtd selectants for binding to lysozyme were expressed as fusion proteins to glutathione S-transferase (GST) before purification and biophysical assays. The wild-type Mtd variant contains the sequence from the original prophage, and provides a negative control for the binding studies. The Mtd-GST fusions were purified by glutathione affinity chromatography to >95% homogeneity, as estimated by SDS-PAGE. The most abundant clone from the BP selections, L-Mtd, demonstrated the highest apparent binding affinity of the over-expressed and purified Mtd variants, as measured by ELISA (Figure 2-3). The dominance of the L-Mtd variant in screens suggests the selection conditions successfully isolated strong binders from the Bvg$^-$-SMPL. Thus, further biophysical characterization focused on L-Mtd.

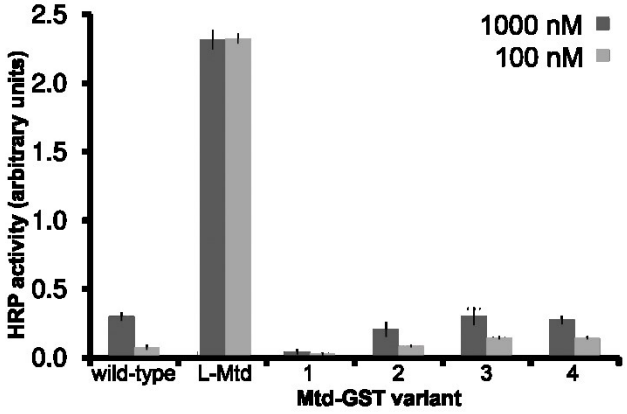


Figure 2-3. The L-Mtd receptor isolated from the BP SMPL binds specifically to T4 lysozyme. In this ELISA, lysozyme has been adsorbed to a microtiter plate at the indicated concentration. L-Mtd expressed expressed as a Mtd-GST fusion protein binds with greater affinity to lysozyme than other selectants, used as controls here, from the library. Error bars indicate standard deviation (n=4).

Size-exclusion chromatography of the wild-type Mtd and L-Mtd fusion proteins demonstrates that the overexpressed and purified fusion proteins form hexamers *in vitro* (Figure 2-4). This hexameric state could result from GST-mediated dimerization of the Mtd trimer, which has also been observed for other proteins[21]. Notably, cryo-electron microscopy imaging with gold-labeled Mtd on the surface of BP also suggests that the Mtd forms a hexamer on the BP tail fibers[12]. Alternatively, if the Mtd is expressed as a trimer on the tail fiber, similar to other *Podoviridae* bacteriophages, the Mtd-GST would still correctly present the Mtd as a homotrimer, but fused to an additional Mtd trimer.

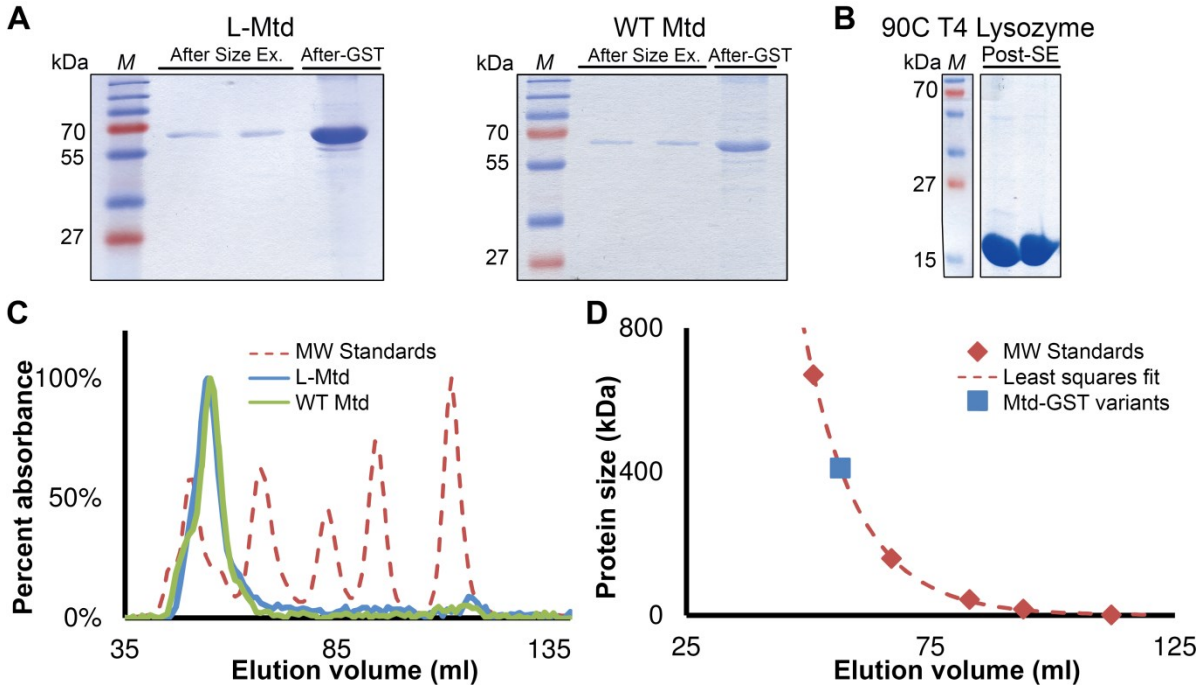
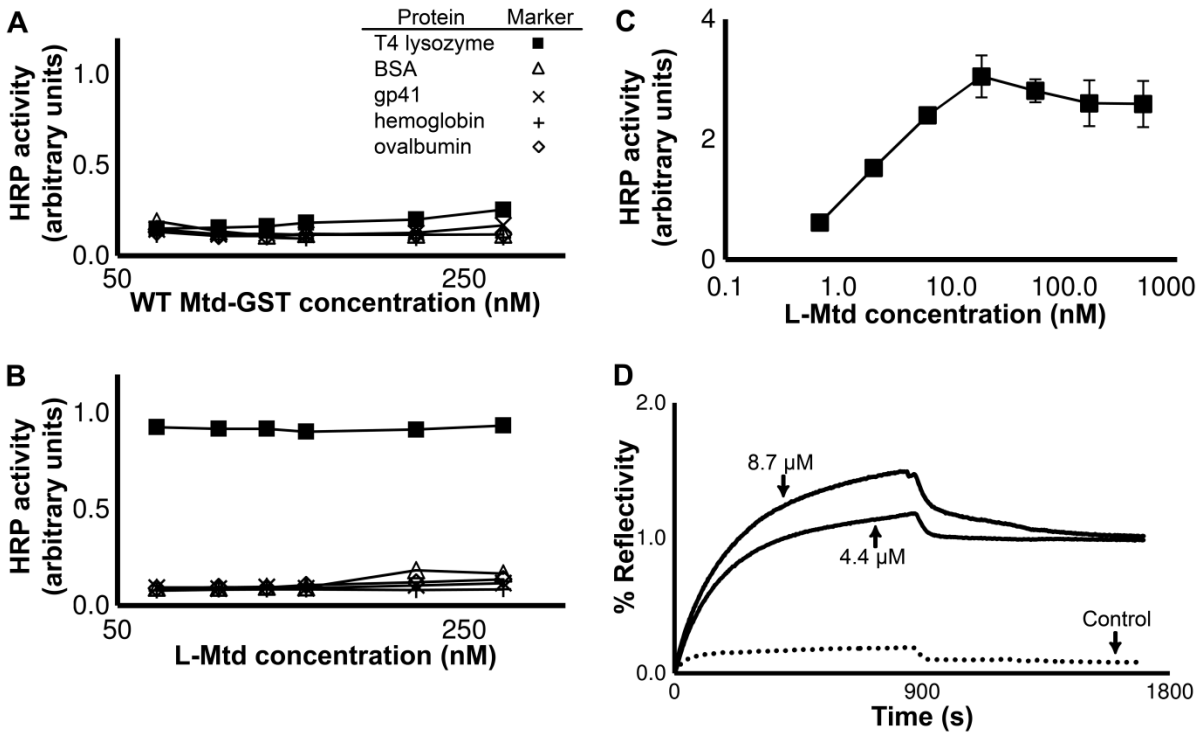


Figure 2-4. Characterization of recombinantly expressed and purified (A) Mtd–GST fusion protein and (B) T4 lysozyme. Mtd-GST variants were purified by GST affinity chromatography followed by size exclusion chromatography (SEC). The T4 lysozyme was purified by cation exchange chromatography. (C) The Mtd-GST fusion proteins were purified by size exclusion chromatography, which was calibrated with MW standards (dashed line, Bio-Rad). (D) A least squares fit to an exponential decay model of the data from size exclusion with MW standards allows estimation of protein size as a function of elution volume. The measured MW of approximately 410 kDa approaches the theoretical size of the Mtd-GST hexamer (406 kDa).

2.3 Characterization of L-Mtd binding to lysozyme

The selectant from the library, L-Mtd, and the negative control, wild-type Mtd, were next examined for binding to lysozyme and additional control proteins. In an ELISA, the wild-type Mtd had no affinity for lysozyme or four control proteins (Figure 2-5A). The L-Mtd variant, however, binds with high affinity to the lysozyme target and does not bind to BSA, HIV gp41, hemoglobin or ovalbumin (Figure 2-5B). This demonstrates the specificity of the L-Mtd selectant. An additional ELISA with decreased

concentrations of lysozyme validates the high affinity interaction between L-Mtd and lysozyme (Figure 2-5C).



	Rate and equilibrium constants
k_{on}	1.15 (±0.20) × 10^3 M^{-1} s^{-1}
k_{off}	5.59 (±0.68) × 10^{-4} s^{-1}
K_D	448 (±67) nM

Figure 2-5. Dose-dependent binding by the recombinantly overexpressed and purified wild-type Mtd and L-Mtd to lysozyme and control proteins. (A) ELISA experiments demonstrate that wild-type Mtd at the indicated concentrations on the X-axis fails to bind to the targeted proteins (coated at 200 nM concentrations) (n=4). (B) L-Mtd at the indicated concentrations on the X-axis, however, binds with high affinity and specificity to lysozyme and not to the other targeted proteins (coated at 200 nM concentrations) (n=4). (C) An ELISA repeated on a separate microtiter plate with lower L-Mtd protein concentrations avoids saturation of HRP activity, and demonstrates the high affinity binding by L-Mtd for lysozyme. Error bars in A, B and C depict standard deviation (n=4). (D) Binding experiments using surface plasmon resonance imaging (SPRI) to measure the affinity of L-Mtd for lysozyme or BSA (negative control). This assay measures the percent change in reflectivity as L-Mtd binds to lysozyme (solid line) or BSA (dotted line) conjugated to the gold layer (solid line). In the table, the rate and

dissociation constants were calculated by a least squares fit to the SPRI data with the standard deviation indicated in parentheses (n=6, single experiment shown).

To further characterize binding affinity, SPRI was used to assay binding by L-Mtd to immobilized lysozyme and BSA (negative control). The target proteins were conjugated covalently to a self-assembled monolayer on the gold surface. The change in percent reflectivity over time allows quantification of the kinetics for L-Mtd binding to either lysozyme or BSA. SPRI measurements demonstrate that the L-Mtd - lysozyme interaction has a dissociation constant, K_D, of 448 ± 67 nM. As expected from the ELISA experiments, L-Mtd does not bind to BSA, the negative control (Figure 2-5D). Interestingly, the L-Mtd to lysozyme interaction slows the off-rate by over two orders of magnitude compared to the off-rate of the wild-type Mtd binding to pertactin, at 5.59×10^{-4} s^{-1} and 3.41×10^{-2} s^{-1}, respectively[13]. Decreasing the off-rate is pivotal to robust protein-protein binding as the intrinsic on-rate constant rarely exceeds 5×10^{6} M^{-1} s^{-1} for protein-protein interactions[22,23].

Presenting multiple ligands in a complex can increase the binding affinity through an avidity effect. Previous experiments have demonstrated that the interaction between the wild-type Mtd and pertactin is increased 10^6-fold when the Mtd is expressed as part of the whole *Bordetella* phage[13]. The weaker monovalent interactions for each Mtd can be amplified by an avidity effect at both the protein scale, with the hexameric Mtd conformation, and at the whole phage scale, with six individual Mtd-containing tail fibers[13]. This potential for multiple levels of avidity could be exploited to obtain large numbers of initial leads, which could be matured into high affinity monomeric receptors. However, the avidity effect requires appropriate geometry between subunits to amplify

the apparent receptor-ligand binding affinity. Further experiments are required to determine if avidity effects contribute to the L-Mtd – lysozyme binding interaction reported here.

2.4 Bordetella bacteriophage as a lead generation system

Tropism switching introduces a large difference in diversity between the naïve Bvg^+- and the Bvg^--SMPLs. Binding a new host receptor to infect the Bvg^- bacteria requires the Bvg^+-derived phage to undergo mutagenesis and tropism switching. Thus, the clones from the naïve Bvg^--SMPL included abundant mutations in the *vr* of the *mtd* (Table 2-1). Furthermore, the high occurrence of tyrosine residues suggests that the diversity accessed by the Bvg^--SMPL is well-suited for identifying new binding partners. Though a high percentage of mutations were observed only in the naïve Bvg^--SMPL, the L-Mtd variant was isolated after selections from both the Bvg^+-SMPL and Bvg^--SMPL libraries. This enrichment ideally reflects the L-Mtd's high affinity for the target, but could also result from cross-contamination or other trivial reason (e.g., much more vigorous growth by BP displaying L-Mtd). Significantly, the L-Mtd variant was never observed in the naïve libraries (Table 2-1), which demonstrates the effectiveness of the selection conditions. The specific, high affinity binding to a novel target demonstrated by the L-Mtd attests to the robustness of the BP-SMPL.

The approach presented here benefits from the high mutation rate and solvent-exposed accessibility of the BP Mtd VR to uncover unique binding partners to target proteins. However, all molecular display systems are subject to important caveats. Conventional phage display, for example, can be finicky in operation. Other

mutagenesis approaches, such as growth in XL1 Red *E. coli* can introduce mutations throughout a plasmid[24]. Generation of the Bvg$^-$-SMPL offers practical advantages by solving the library synthesis challenge and offering a highly diverse, tyrosine-rich library targeted to a single region of a specific ORF. Growth of the *B. bronchiseptica* Bvg$^+$ bacteria requires Biosafety Level 2 precautions, analogous to tissue culture experiments. We recommend working with the avirulent Bvg$^-$ bacteria, which can be used outside a tissue culture hood. As demonstrated here the Bvg$^-$-SMPL obtained from the avirulent strain also has greater sequence diversity for selections than the Bvg$^+$-SMPL. However, mutagenesis is based on phage propagation, and multiple rounds of selection are not possible unlike conventional M13-based libraries. In addition, adenine substitution of the AAC, ACG, and ATC codons in the native *mtd tr* does not allow mutagenesis to codons encoding the amino acids glutamate, glutamine, lysine, methionine, and tryptophan[18].

Engineered, non-immunoglobin protein scaffolds draw inspiration from antibody-based molecular recognition[25–29]. For example, like antibodies, the BP-SMPL provides malleable loops on the surface of the protein, which can be adapted to recognize different binding partners; such loops are analogous to the CDR regions of antibodies. In addition to its self-synthesis of protein libraries, the BP-SMPL includes other desirable attributes for a non-immunoglobulin protein scaffold, such as high protein yields from over-expression. Binding ligands identified from selections with the BP-SMPL could be useful in biosensor and single-molecule studies[15,16,30,31].

The dominance of tyrosine residues in the Mtd selectants reported here suggests that results obtained with minimally substituted antibody libraries could be generalized

to other adaptable binding scaffolds[19,32]. This bias for tyrosine substitutions likely increased the potential for obtaining strongly binding receptors. Thus, selections with the tyrosine-rich SMPL from Bvg$^-$ can reduce the time and costs associated with library-based selections. Additionally, prior experiments to enable endogenous mutagenesis of genes heterologous to the *mtd vr* suggest that *Bordetella* phage can expedite directed evolution of a wide variety of protein targets[8,33]. The high mutation rates generated within the *mtd vr* and the ability to modify adenine-encoded mutational hot spots provides a system for automated mutagenesis of introduced foreign sequences. The approach could overcome the diversity barriers set by the efficiency of bacterial transformation, which is inherent to conventional phage display systems.

2.5 Materials and Methods

2.5.1 Generation of BP-SMPLs

To acquire sufficient phage for selections, four cycles of infection were completed to generate the two BP-SMPLs. a single colony of Bvg$^+$ phase bacteria containing the BP prophage was cultured in 3 mL of 2x LB (20 g tryptone, 10 g yeast extract, 5 g NaCl, 1 l H_2O, pH 7.0) for 16 h at 37 °C and shaken at 225 rpm. The culture was transferred to two Eppendorf tubes, and centrifuged at 12 krpm for 5 min at room temperature. The supernatant was filtered (0.2 µm) to obtain pure BP. The following soft-overlay method was used to isolate viable phage[34]. Briefly, 115 µl of the filtered supernatant and 230 µl of log-phase Bvg$^+$ phase bacteria were added to 3 mL of 42 °C 0.7% w/v top agar (0.7 g agar, 100 mL LB). The mixture was vortexed briefly and then poured over a pre-warmed 115x10 mm LB agar plate (10 g tryptone, 5 g yeast extract, 10 g NaCl, 15 g agarose, 1

L H_2O) containing streptomycin (40 μg mL^{-1}) before incubating for 16 h at 37 °C. BP particles were removed from the translucent plate by adding 3 mL of SM buffer (5.8 g NaCl, 2 g $MgSO_4 \cdot 7\ H_2O$, 5 mL 2% w/v gelatin, 1 L H_2O) followed by incubation at 4 °C for 3 h on a rotary shaker with gentle rocking. After incubation, the supernatant containing viable phage was removed from the plates and filtered. The soft overlay method using 50 μl of either Bvg^+ or Bvg^- phase bacterial cells was used to generate titer plates supporting a lawn of bacteria on LB plates containing streptomycin (40 μg mL^{-1}). The supernatant containing viable phage was diluted in 10-fold dilutions in SM buffer and titered on LB-streptomycin plates. Plates were incubated for 18 h at 37 °C, and the plaque forming units counted.

Each propagation cycle was repeated using the filtered phage from the previous cycle. The second propagation in Bvg^+ phase bacteria further increased phage titers. For the third propagation, the filtered BP were divided into two aliquots to infect either Bvg^+ phase bacteria or Bvg^- phase bacteria. The two separate SMPLs were re-propagated one additional time in the respective host bacteria to boost phage titers (Figure 2-1C). Final SMPLs were normalized to 1.2×10^8 PFU mL^{-1} by dilutions in SM buffer.

2.5.2 Selections with SMPLs

Selections for binding to lysozyme isolated members of the BP-SMPL with low off-rates, using an adaption of previous phage display experiments[35]. Experiments with lysozyme used a pseudo-wild-type variant of lysozyme (a gift from Brian Matthews, University of Oregon) with the following mutations designed to introduce a single cysteine residue into the protein: C54T, S90C, and C97A. Following overexpression and purification of

the lysozyme variant by cation exchange chromatography (linear gradient, 0.01 to 1 M NaCl, 20 mM Tris, pH 7.5), the protein was incubated in the reducing agent tris(2-carboxyethyl)phosphine hydrochloride (200 µM), followed by covalent attachment to biotin through incubation with maleimide-PEG_2-biotin (0.5 mM, Thermo Scientific) for 2 h on ice[36]. The biotinylated lysozyme was then dialyzed into PBS with 0.2% BSA and 0.05% Tween 20 before incubation with phage libraries from the Bvg^+-SMPL and Bvg^--SMPL for 1 h. Next, 100 molar excess non-biotinylated lysozyme (5 µM) was added to the lysozyme - phage library solution (50 nM) for 40 minutes to remove weakly bound phage variants. Capture of the lysozyme-bound phage was accomplished by adding streptavidin-coated paramagnetic beads (0.5 mg, Invitrogen) to the solution. The beads were incubated for five min before rinsing seven times with wash buffer (PBS, 0.2% BSA, 0.05% Tween 20). Phage that remained bound to the magnetic beads were amplified by PCR for DNA sequencing. DNA sequences and primers have been submitted to GenBank (accession number pending, submission ID: 1590838).

2.5.3 Mtd gene isolation and expression of Mtd-GST

After selection, the *mtd* genes of the phage were amplified by standard PCR protocols using primers with encoded *BamHI* and *XhoI* restriction sites (Table 2-3). Standard sequencing methods were used to prepare DNA for sequencing by the Genewiz DNA Sequencing Service. PCR amplicons of the *mtd* gene were ligated into the pGEX-6P-3 expression vector (GE Healthcare). PCR was used to amplify 3 µl of eluted phage in a 28 µl reaction (17 µl H_2O, 5.6 µl 5x FlexiBuffer buffer (Promega), 1.5 µl 50 mM $MgCl_2$, 0.25 µl 25 mM dNTPs, 0.3 µl (5U/µl) GoTaq polymerase (Promega),

0.25 µl 330 ng/µl *VR-Fwd*, 0.25 µl 330 ng/µl *VR-Rev*). The reaction was cycled in a thermocycler (95 °C for 5 min, 38 x (95 °C for 1 min, 65 °C for 30 s, 72 °C for 1 min), and then 72 °C for 7 min). The resultant plasmid was sub-cloned into a heat shock competent *E. coli* strain (Top Ten) before rescue by addition of SOC media (2% tryptone, 0.5% yeast extract, 8.56 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 20 mM glucose) for 50 min at 37 °C. The rescued *E. coli* were then spread on LB plates supplemented with carbenicillin (50 µg/ml)[37].

Primer Table

Name	*Sequence*	*Restriction sites*
GstMtdBamHTruncFwd	CGC GGA TCC AGT ACC GCA GTC CAG TTC CGC	BamH1
MtdMutXhoIRev	GAC CTC GAG TCA CTA CTC AAG AAT CAG GTG GTC ACA GAC	XhoI
VR-Fwd	TGT AAA ACG ACG GCC AGT TAG CAC TTT GTC GCT TCC	-
VR-Rev	CAG GAA ACA GCT ATG ACT GGC GCA TCC GAA TAC AC	-

Table 2-3. Primers utilized for *vr* and *mtd* gene sequencing. Restriction sites were added to allow sub-cloning into the pGEX-6P-3 vector (GE Healthcare).

2.5.4 Mtd-GST protein purification

Mtd variants were transformed and overexpressed in BL21 *E. coli* bacteria as Mtd-GST fusion proteins after induction with isopropyl β-D-1-thiogalactopyranoside (IPTG, 1 mM) for 16 h at 22 °C with shaking at 150 rpm. Cell lysate from a 1 L culture was centrifuged at 5000 rpm (3468 x g) for 30 min, at 10 °C. The cell pellet was reconstituted in 20 ml lysis buffer (500 mM NaCl, 50 mM phosphate, 10 mM imidazole, 1 mM Halt protease inhibitor from Pierce, 10 mM 2-mercaptoethanol, pH 8.0), and sonicated in 30 s continuous bursts with 1 min cooling times for six cycles (20 watts). The sonicated solution was then centrifuged, and the supernatant was applied to an equilibrated GST-Bind resin (3 ml bed volume, Novagen) for 30 min at room

temperature. The column was washed with five bed volumes of GST wash buffer (43 mM Na_2HPO_4, 14.7 mM KH_2PO_4, 1.37 M NaCl, 27 mM KCl, pH 7.3). The protein was eluted with glutathione reconstitution buffer (500 mM Tris-HCl, 10 mM reduced glutathione, pH 8.0). The cell pellet and soluble fractions were analyzed by SDS-PAGE (12% acrylamide).

Mtd-GST protein and lysozyme were further purified by size exclusion chromatography according to manufacture protocols following dialysis into PBS with 3000 MWCO dialysis tubing (BioRad). Overnight cultures of 1 L routinely yielded over 8 mg of purified protein for both wild-type and L-Mtd variants (Table 2-4).

Variant	*mg/ml*	*Volume (ml)*	*Yield (mg)*
Wild-type Mtd	6.327	1.50	9.49
L-Mtd	5.541	1.50	8.31

Table 2-4. Protein yields for 1 L of overexpressed wild-type Mtd and L-Mtd after purification by GST affinity chromatography and size exclusion chromatography.

2.5.5 ELISA binding studies

ELISAs were used to examine binding to lysozyme, BSA, hemoglobin, gp41, and ovalbumin by purified, recombinant L-Mtd, wild-type Mtd, and four additional Mtd variants selected for binding to lysozyme. Maxisorp 96-well microtiter plates were coated with 200 µl of the target protein solution (200 nM, unless otherwise noted, in 50 mM $NaCHO_3$, pH 9.6), blocked with 0.2% non-fat milk (NFM) in PBS, rinsed with wash buffer and incubated with the Mtd-GST fusion protein variants (1 µM in blocking buffer) for 1 h. The microtiter plates were then incubated with polyclonal rabbit anti-Mtd antibody (1:500) and HRP-conjugated anti-rabbit antibody (1:5000) in blocking buffer for

1 h each. The ELISA was developed by the addition of 1% w/v *o*-phenylenediamine dihydrocholoride in citric acid buffer (0.02% w/v H_2O_2, 50 mM citric acid, 50 mM Na_2HPO_4, pH 5.0), and the absorbance was measured at 450 nM using a microtiter plate reader.

2.5.6 Surface plasmon resonance biosensor experiments

Surface plasmon resonance imaging (SPRI) experiments were performed as previously described using a SPRimager II biosensor instrument with SpotReady pure gold sensor chips (GWC Technologies)[38]. Protein targets (lysozyme and BSA at 10 μM) were immobilized on the 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide/*N*-hydroxysuccinimide treated sensor chips (0.3 μl in PBS, 1 mM $MgCl_2$, pH 8.0) for 1 h. The sensor chip was then blocked with 0.5% NFM in the same buffer (0.8 μl). All measurements were preceded by conditioning the sensor chip with flowing buffer for 30 s. One ml of Mtd-GST solution (8.7 or 4.4 μM in running buffer) was flowed across the sensor chip at 1 μl/s. The increase in pixel intensity for each spot was averaged at 30 frames per s for 900 s. All measurements were conducted in triplicate with subtraction of the background measured for spots treated only with blocking buffer.

2.6 Acknowledgements

The authors thank Prof. J. Miller, A. Hodes and S. Doulatov (UCLA) for supplying wild-type bacteria, bacteriophage, and anti-phage antibodies used in these studies and Dr. S. Majumdar for his technical expertise. This work was financially supported by the National Institute of General Medical Sciences of the NIH (R01 GM078528-01).

REFERENCES

1. Scott, J. K. & Smith, G. P. Searching for peptide ligands with an epitope library. *Science.* **249,** 386–390 (1990).

2. Kehoe, J. W. & Kay, B. K. Filamentous phage display in the new millennium. *Chem. Rev.* **105,** 4056–4072 (2005).

3. Sidhu, S. S. Phage display in pharmaceutical biotechnology. *Curr. Opin. Biotechnol.* **11,** 610–616 (2000).

4. Smothers, J. F., Henikoff, S. & Carter, P. Affinity selection from biological libraries. *Science.* **298,** 621–622 (2002).

5. Doulatov, S. *et al.* Tropism switching in Bordetella bacteriophage defines a family of diversity-generating retroelements. *Nature.* **431,** 476–481 (2004).

6. Overstreet, C. M. *et al.* Self-made phage libraries with heterologous inserts in the Mtd of Bordetella bronchiseptica. *Protein Eng. Des. Sel.* **25,** 145–151 (2012).

7. Liu, M. *et al.* Genomic and genetic analysis of Bordetella bacteriophages encoding reverse transcriptase-mediated tropism-switching cassettes. *J. Bacteriol.* **186,** 1503–1517 (2004).

8. Akerley, B. J., Cotter, P. A. & Miller, J. F. Ectopic expression of the flagellar regulon alters development of the Bordetella-host interaction. *Cell.* **80,** 611–620 (1995).

9. Uhl, M. A. & Miller, J. F. Integration of multiple domains in a two-component sensor protein: the Bordetella pertussis BvgAS phosphorelay. *EMBO J.* **15,** 1028–1036 (1996).

10. Bock, A. & Gross, R. The BvgAS two-component system of Bordetella spp.: a versatile modulator of virulence gene expression. *Int. J. Med. Microbiol.* **291,** 119–130 (2001).

11. Liu, M. *et al.* Reverse transcriptase-mediated tropism switching in Bordetella bacteriophage. *Science.* **295,** 2091–2094 (2002).

12. Dai, W. *et al.* Three-dimensional structure of tropism-switching Bordetella bacteriophage. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 4347–4352 (2010).

13. Miller, J. F. J. L. F., Le Coq, J., Hodes, A., Barbalat, R. & Ghosh, P. Selective ligand recognition by a diversity-generating retroelement variable protein. *PLoS Biol.* **6,** e131 (2008).

14. Medhekar, B. & Miller, J. F. Diversity-generating retroelements. *Curr. Opin. Microbiol.* **10,** 388–395 (2007).

15. Choi, Y. *et al.* Single-molecule dynamics of lysozyme processing distinguishes linear and cross-linked peptidoglycan substrates. *J. Am. Chem. Soc.* **134,** 2032–2035 (2012).

16. Choi, Y. *et al.* Single-molecule lysozyme dynamics monitored by an electronic circuit. *Science.* **335,** 319–24 (2012).

17. Mian, I. S., Bradwell, A. R. & Olson, A. J. Structure, function and properties of antibody binding sites. *J. Mol. Biol.* **217,** 133–151 (1991).

18. McMahon, S. A. *et al.* The C-type lectin fold as an evolutionary solution for massive sequence variation. *Nat. Struct. Mol. Biol.* **12,** 886–892 (2005).

19. Koide, S. & Sidhu, S. S. The Importance of Being Tyrosine: Lessons in Molecular Recognition from Minimalist Synthetic Binding Proteins. *ACS Chem. Biol.* **4,** 325–334 (2009).

20. Fellouse, F. A. *et al.* High-throughput Generation of Synthetic Antibodies from Highly Functional Minimalist Phage-displayed Libraries. *J. Mol. Biol.* **373,** 924–940 (2007).

21. Dixon, D. P., Lapthorn, A. & Edwards, R. Plant glutathione transferases. *Genome Biol.* **3,** 3004.1–3004.10 (2002).

22. Northrup, S. H. & Erickson, H. P. Kinetics of protein-protein association explained by Brownian dynamics computer simulation. *Proc. Natl. Acad. Sci. U. S. A.* **89,** 3338–3342 (1992).

23. Schlosshauer, M. & Baker, D. Realistic protein-protein association rates from a simple diffusional model neglecting long-range interactions, free energy barriers, and landscape ruggedness. *Protein. Sci.* **13,** 1660–1669 (2004).

24. Coia, G., Ayres, A., Lilley, G. G., Hudson, P. J. & Irving, R. A. Use of mutator cells as a means for increasing production levels of a recombinant antibody directed against Hepatitis B. *Gene* **201,** 203–209 (1997).

25. Gebauer, M. & Skerra, A. Engineered protein scaffolds as next-generation antibody therapeutics. *Curr. Opin. Chem. Biol.* **13,** 245–255 (2009).

26. Ebersbach, H. *et al.* Affilin-novel binding molecules based on human gamma-B-crystallin, an all beta-sheet protein. *J. Mol. Biol.* **372,** 172–185 (2007).

27. Binz, H. K., Amstutz, P., Plückthun, A. & Pluckthun, A. Engineering novel binding proteins from nonimmunoglobulin domains. *Nat. Biotechnol.* **23,** 1257–68 (2005).

28. Hosse, R. J., Rothe, A. & Power, B. E. A new generation of protein display scaffolds for molecular recognition. *Protein Sci.* **15,** 14–27 (2006).

29. Roberts, B. L. *et al.* Directed evolution of a protein: selection of potent neutrophil elastase inhibitors displayed on M13 fusion phage. *Proc. Natl. Acad. Sci. U. S. A.* **89,** 2429–2433 (1992).

30. Arter, J. a *et al.* Virus-polymer hybrid nanowires tailored to detect prostate-specific membrane antigen. *Anal. Chem.* **84,** 2776–2783 (2012).

31. Arter, J. A., Taggart, D. K., McIntire, T. M., Penner, R. M. & Weiss, G. A. Virus-PEDOT nanowires for biosensing. *Nano Lett.* (2010). doi:10.1021/nl1025826

32. Sidhu, S. S. & Kossiakoff, A. a. Exploring and designing protein function with restricted diversity. *Curr. Opin. Chem. Biol.* **11,** 347–354 (2007).

33. Guo, H. *et al.* Target site recognition by a diversity-generating retroelement. *PLoS Genet* **7,** e1002414 (2011).

34. Adams, M. H. *Bacteriophages*. (Interscience Publishers, Inc., 1959).

35. Hawkins, R. E., Russell, S. J. & Winter, G. Selection of Phage Antibodies by Binding Affinity Mimicking Affinity Maturation. *J. Mol. Biol.* **226,** 889–896 (1992).

36. Han, J. C. & Han, G. Y. A procedure for quantitative determination of tris(2-carboxyethyl)phosphine, an odorless reducing agent more stable and effective than dithiothreitol. *Anal. Biochem.* **220,** 5–10 (1994).

37. Chung, C. T., Niemela, S. L. & Miller, R. H. One-step preparation of competent Escherichia coli: transformation and storage of bacterial cells in the same solution. *Proc. Natl. Acad. Sci. U. S. A.* **86,** 2172–2175 (1989).

38. Subramanian, L., Walker, T. M., Takach, J. C. & Polans, A. S. An improved surface plasmon resonance method for determining the molecular basis of the calcium-dependent interactions of ALG-2. *Calcium Bind. Proteins* **1,** 54–61 (2006).

CHAPTER 3

Shear stress-mediated refolding of proteins from aggregates and inclusion bodies

Abstract

Recombinant protein overexpression of large proteins in bacteria typically results in insoluble and misfolded proteins directed to inclusion bodies. We report the application of shear stress in micrometer-wide, thin fluid films to mimic two chaperonin mechanisms – forced unfolding of the substrate protein and segregation away from misfolded intermediates. Shear stress successfully refolded boiled hen egg white lysozyme, recombinant hen egg white lysozyme, and recombinant caveolin-1. Furthermore, site-specific shear stress allowed refolding of the much larger protein, cAMP-dependent protein kinase A (PKA). The reported methods require only minutes, which is >100-times faster than conventional, overnight dialysis. This rapid refolding technique could significantly shorten the times, lower costs, and reduce the waste streams associated with protein expression for a wide-range of industrial and research applications.

Introduction

Overexpressed recombinant proteins for industrial, pharmaceutical, environmental and agricultural applications annually represent a >$160 billion biotechnology world market.[1] Protein expression in yeast or *Escherichia coli* is highly preferred due to the organisms' rapid growth, low consumable costs, and high yields[2,3]. However, large proteins overexpressed in bacteria typically form aggregates and inclusion bodies[4–7]. Recovery of the correctly folded protein then requires laborious and expensive processing of inclusion bodies by conventional methods[1,8]. The most common method for refolding such proteins, for example, involves multi-day dialysis with large volumes (typically 1-10 liters for mg quantities of protein)[9].

Alternatively, high value proteins (e.g., therapeutic antibodies or GPCRs for structural biology) apply extensively optimized mammalian or insect cell lines, media and bioreactor conditions[10–12]. Recovery of correctly folded proteins from aggregates is inefficient and challenging for large-scale industrial processes. One method to solve this problem applies very high hydrostatic pressures (400 bar) to refold recombinant proteins from inclusion bodies[13,14]. New methods capable of broadening the utility of bacterial over-expression could transform industrial and research production of proteins.

3.1 Strategies for protein refolding

Nature has evolved molecular machines, termed chaperones, to assist with protein folding. One class of these machines, chaperonins (e.g., GroEL-GroES in *E. coli*), can reverse small protein aggregates, and refold proteins (Figure 3-1A)[15–20]. This assistance is required by essentially all proteins >100 residues in length produced in

cells[21]. After binding to the substrate protein, ATP hydrolysis by GroEL triggers unfolding of the misfolded protein[19]; during this step, the chaperonin undergoes a conformational rearrangement to unfold the protein. Then, ATP-dependent binding of the GroES complex allows the protein to refold in ≈10 s while enclosed in a hydrophilic, cage-like interior of the chaperonin[17,20,22]. Thus, the GroEL-GroES chaperonin system embodies two important refolding concepts – mechanical unfolding and shielding of partially folded intermediates. Inspired by the chaperonin's mechanism, we demonstrate *in vitro* renaturation of denatured proteins by mechanically driven unfolding and separation from other intermediates during refolding.

3.1.1 Characterizing shear forces in the vortex fluid device

We report using a vortex fluid device (VFD) to apply shear forces for rapid equilibration of protein folding and isolation of intermediates during protein folding. In this method, a glass cylinder (10 mm by 16 cm) is spun rapidly (5 krpm) at a 45° angle. At high rotational speeds, the solution within the sample tube forms micrometer-thick, thin fluid films, which flow with the same speed and direction as the wall of the glass tube. The rotating glass tube generates a velocity gradient within the thin fluid film, which introduces shear stress into the solution (Figure 3-1B). Application and optimization of a similar vortex fluid device was first demonstrated for exfoliating graphite and hexagonal boron-nitride to generate mono and multi-layer structures[23,24]. More recently,the approach has been used to control the formation of different calcium carbonate polymorphs[25].

3.1.2 Modeling fluid dynamics in thin fluid films

Modeling the fluid behavior in the VFD allowed estimation of the shear forces experienced by proteins folding at various rotational speeds. Our analysis applies the solution for cylindrical Couette flow[26,27]. The velocity of the solution, v_θ, is a function of the radius, r (Figure 3-1B). The boundary conditions for the liquid film interfaces are defined as follows. The inner air-liquid interface at $r = R_1$ slips due to discontinuity in viscosity, and results in vanishingly low shear stress ($\frac{dv_\theta}{dr} = 0$). At the outer liquid-glass interface, the no-slip boundary dictates that the velocity of the liquid at $r = R_2$ matches that of the inner wall of the glass tube ($v_\theta = R_2 \cdot \Omega$), where Ω is the angular velocity of the tube. The resulting velocity profile is a nonlinear function of the form

$$v_\theta = Ar + \frac{B}{r}$$

where

$$A = \frac{\Omega}{\frac{R_1^2}{R_2^2} + 1} \quad \text{and} \quad B = \frac{\Omega R_1^2}{\frac{R_1^2}{R_2^2} + 1}.$$

From this velocity profile, shear stress can be calculated as

$$\tau_{r\theta} = \mu r \frac{\partial}{\partial r}\left(\frac{v_\theta}{r}\right)$$

where μ is the viscosity of water at 20 °C. At a speed of 5 krpm, the calculated shear stress ranges from 0.53 to 0.56 Pa (Figure 3-1C). These values of shear stress are similar to the requirements previously reported for protein unfolding[28].

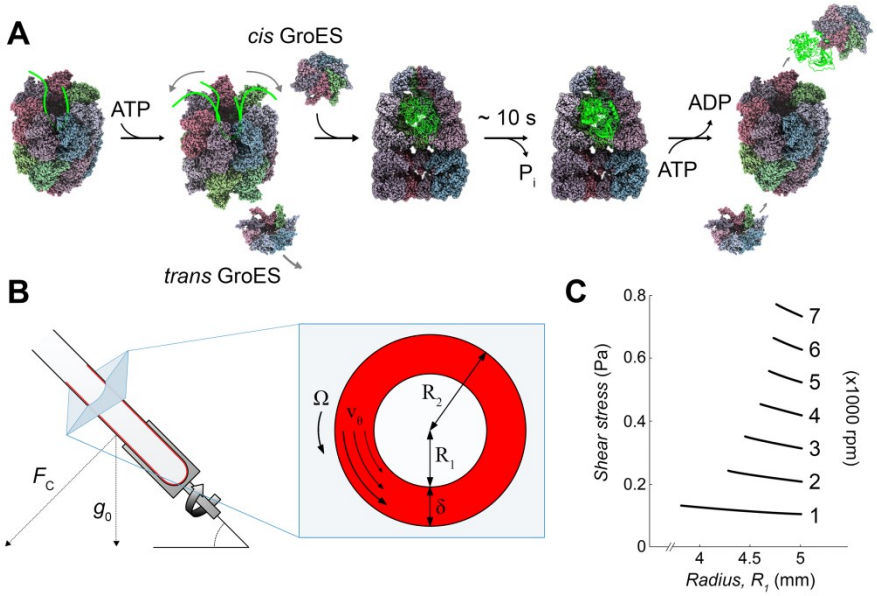


Figure 3-1. Protein refolding *in vitro* with the vortex fluid device (VFD) generates shear flow inside thin fluid films. (A) Protein substrate undergoes forced unfolding upon ATP binding to the *trans* GroEL-GroES complex. Unfolding allows the protein to reach higher energy conformers prior to encapsulation within the *cis* GroEL-GroES complex (PDB: 1AON, 4AB2). (B) Modeling movement in the thin fluid film. The inset represents a two-dimensional slice of the tube along the axis of rotation. The radius of the tube, R_2, is 5 mm, and the thin fluid film thickness, δ, has been previously measured (not to scale)[23]. The velocity of the solution, v_θ, increases from the inner surface of the film to match the velocity of the tube, Ω. (C) Shear stress calculated as a function of rotational speed (rpm) and radius from the center of the tube.

3.2 Refolding denatured hen egg white lysozyme from egg whites

Experiments with native hen egg white were conducted to determine if shear forces could refold denatured hen egg white lysozyme (HEWL) in complex environments. The separated whites were diluted in PBS, and heat-treated at 90 °C for 20 min. The resultant hard-boiled egg white was dissolved in 8 M urea, rapidly diluted and then VFD processed at the indicated rotational speeds and times (Figure 3-3). Total protein concentration as determined by bicinchoninic acid assay was 44 µg/ml. Recovery of HEWL activity was then demonstrated by a lysozyme activity assay (Figure

3-2). Refolding HEWL within the egg white at 5 krpm recovers activity even after a short 2.5 min spin, but continued shear forces unfolds the protein. VFD processing for 5 minutes at 5 krpm results in optimal HEWL refolding (Figure 3-3A). This experiment establishes parameters for lysozyme refolding by VFD.

Standard	
U	RFU/min
25	0.201
12.5	0.163
6.25	0.071
3.125	0.040
1.5625	0.019
0.78125	0.017
0.390625	0.013
0	0.012

Michaelis-Menten + C	
Best-fit values	
Vmax	0.4187
Km	25.92
C	0.003201
Vmax	0.128
Km	14.46
C	0.008513
Degrees of Freedom	5
R square	0.9735
Absolute Sum of Squares	0.001024

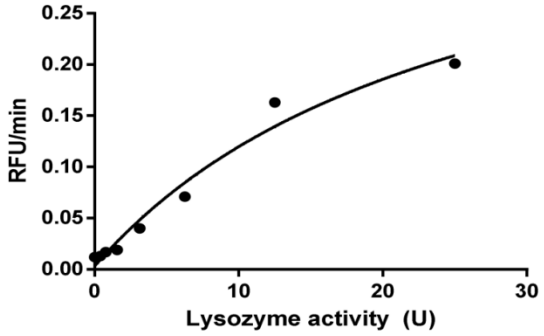


Figure 3-2. Wild-type lysozyme activity interpolated by least-squares regression fit. Relative fluorescence units per minute was fit to standardized lysozyme activity assayed using the EnzChek Lysozyme Activity kit (Invitrogen) to a Michaelis-Menten model, plus background with Prism 6 software (GraphPad). All assays were conducted in 96-well black opaque microtiter plates, 100 µL reaction volumes, with 10 min incubation at 37 °C.

3.3 Refolding recombinant hen egg white lysozyme from inclusion bodies

To demonstrate refolding of recombinantly expressed, reduced HEWL, the cell pellet was reconstitued in lysis buffer containing 2-mercaptoethanol, purified, urea-denatured and rapidly diluted into PBS (1:100, see Table 3-1). Second, the diluted protein (1 ml, 44 µg/ml) was immediately transferred to the VFD sample tube and spun at 22 °C and 5 krpm for 5 min. Circular dichroism (CD) spectra of the VFD-refolded,

recombinant HEWL demonstrates restoration of secondary structure from proteins isolated from inclusion bodies. After VFD processing, the CD spectra of identical HEWL samples demonstrates partial recovery of secondary structure compared to the native lysozyme (Figure 3-3B). Yields determined by protein activity assay are provided in Figure 3-3D.

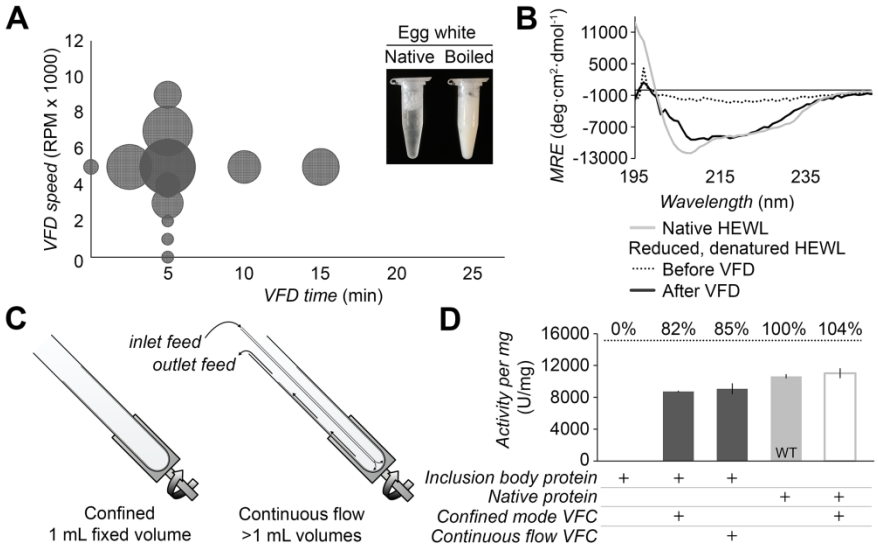


Figure 3-3. Determination of secondary structure and activity of hen egg white lysozyme (HEWL) processed by VFD. (A) Lysozyme activity per mg protein following VFD processing of boiled egg white (90 °C, 20 min) at fixed VFD speed of 5 krpm, or fixed 5 min refolding time. Relative circle size represents the lysozyme activity, plotted as a function of VFD time and speed. (B) CD spectra of recombinantly expressed, reduced, and denatured HEWL before (dotted) and after (dark gray) VFD refolding. (C) In continuous flow mode, the protein solution is introduced through a thin, hollow metal tube to the bottom of the sample tube. (D) Lysozyme activity per mg protein following VFD refolding of recombinantly expressed HEWL. Under these conditions, VFD treatment of purified, active lysozyme does not adversely affect its activity. Values above dotted line represent percent recovery of activity compared to wild-type. Error bars indicate standard deviation (n = 3).

HEWL can also be refolded by continuous flow VFD. This approach delivers additional sample through an inlet at the cylinder base. The sample (50 ml), added at a

flow-rate of 0.1 ml/min, demonstrates significant recovery of HEWL activity for scalable, high volume applications (Figure 3-3C). The recombinant HEWL recovers >82% of its activity following VFD treatment. As expected, HEWL isolated from inclusion bodies without VFD processing fails to show any lysozyme activity (Figure 3-3D). The continuous flow approach could be scaled up for industrial applications.

3.4 Applying shear stress to refold recombinantly expressed caveolin-ΔTM

After refolding denatured lysozyme in both complex (egg white) and simple (purified recombinant protein) environments, the next experiments focused on refolding the protein caveolin-1, as an example of a protein requiring an inordinate amount of processing time by conventional approaches (e.g., four days of dialysis). A caveolin variant without its transmembrane domain (caveolin-ΔTM) was recombinantly expressed, and the inclusion body was purified under denaturing conditions. Purified caveolin-ΔTM was diluted, and then given a short dialysis for 1 h to lower the urea concentration. The protein was then VFD-treated for 0, 10, or 30 min at 5 krpm at a concentration of 186 μg/ml. The CD spectra of the VFD processed caveolin-ΔTM shows a pronounced minima at 208 nm, which is indicative of α-helical secondary structure (Figure 3-4A)[29]. Solution turbidity also decreases sharply following VFD treatment, which illustrates VFD solubilization and refolding of partially aggregated proteins (Figure 3-4B). ELISA experiments examined binding by the refolded caveolin-ΔTM to HIV glycoprotein 41 (gp41), a known caveolin binding partner[30,31]. VFD processing significantly restores protein function, as shown through binding to gp41 (Figure 3-4C).

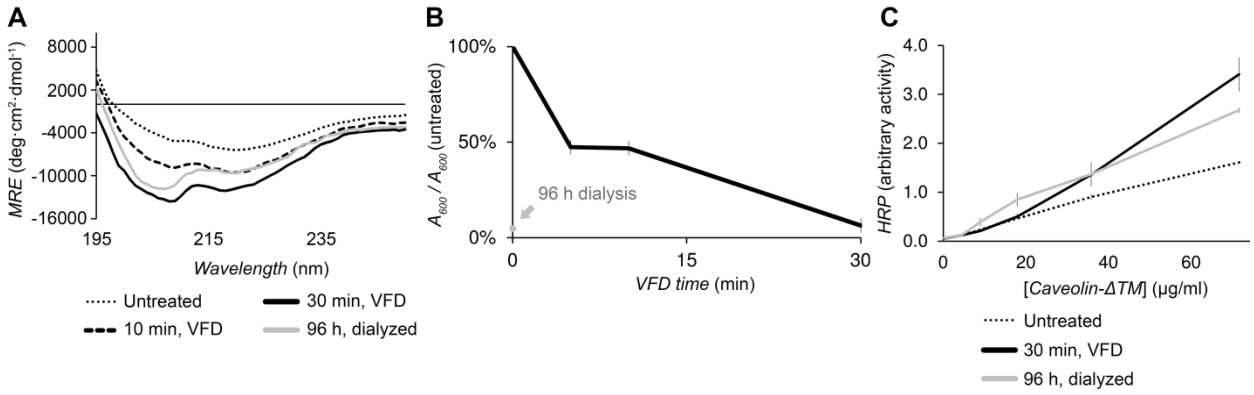


Figure 3-4. Determination of secondary structure and activity of caveolin-ΔTM processed by VFD. (A) Circular dichroism (CD) spectra of caveolin-ΔTM following VFD induced refolding or conventional dialysis. (B) Turbidity of caveolin-ΔTM, measured by the fractional absorbance at 600 nm compared to the untreated sample also at 600 nm. The arrow indicates absorbance of caveolin-ΔTM following 4-days of dialysis. (C) Dose-dependent binding of unprocessed, processed, and conventionally refolded caveolin-ΔTM to HIV gp41 determined by ELISA. Although some binding occurs without VFD treatment, caveolin-ΔTM binds with greater affinity after VFD refolding. Error bars indicate standard deviation (n = 3).

3.5 Recovering kinase activity from PKA by attachment to IMAC beads

Larger-sized proteins initially failed to refold despite VFD treatment. For example, the catalytic domain of PKA (42 kDa) is significantly bulkier than HEWL (14 kDa) and caveolin-ΔTM (17 kDa), and did not refold from inclusion bodies after treatment following similar protocols. To refold full-length PKA *in vitro*, we hypothesized that a closer mimic of cellular folding was required. In cells, the nascent polypeptide can fold as the *N*-terminus extrudes from the ribosome, whereas *in vitro* refolding must address the entire protein at once[32]. Thus, we focused shear stress on the *N*-terminus of His-tagged PKA by immobilization on Ni^{2+}-charged immobilized metal affinity chromatography (IMAC) beads. The IMAC-His-PKA complex was then subjected to shear stress in the VFD (1 ml, 0.2 – 1 mg/ml). Following VFD treatment, His-PKA separated from the IMAC resin, and recovered 69% of its kinase activity (Figure 3-5A).

Negative control experiments with identical conditions, but with uncharged IMAC resin, lower charged resin volumes, or 500 mM imidazole to block the Ni^{2+}-His_6 tag interaction, showed only low levels of kinase activity (Figure 3-5A). Interestingly, the remaining His-PKA eluted from the IMAC resin by elution with imidazole yielded far less active protein (Figure 3-5B).

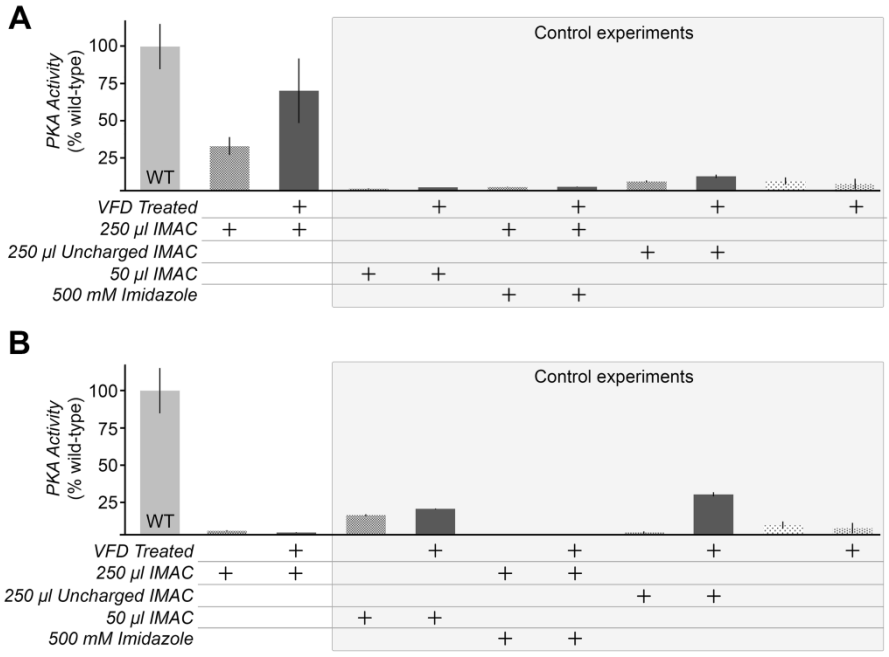


Figure 3-5. Facilitating VFD refolding of PKA by pre-binding to IMAC resin. 1.7 mg PKA was pre-incubated with 250 or 50 µl IMAC resin in 6 M guanidine-HCl prior to dilution to 1 M guanidine-HCl and VFD treatment. (A) PKA is removed by a low imidazole (1 mM) wash buffer. PKA activity per µg of protein measured by nicotinamide adenine dinucleotide (NADH) enzyme-linked assay and shown as percentage of wild-type PKA. Consumption of ATP by active PKA results in consumption of NADH by lactate dehydrogenase. The NADH levels are monitored in this assay by absorbance at 340 nm. Negative control experiments were performed with low quantities of resin, 500 mM imidazole, or resin lacking Ni2+ (uncharged) during IMAC incubation. (B) PKA activity per µg of protein of elution fractions following VFD treatment. Steps performed on elution fractions were identical to wash fractions. Error bars indicate standard deviation (n = 3).

As reported here, protein refolding by VFD requires optimization for each protein. Buffers, protein concentration, and processing time were optimized for HEWL, caveolin-ΔTM, and PKA. The refolding of HEWL from the complex mixture of boiled egg whites appears less efficient than recovery of the folded protein from inclusion bodies. In egg whites, the mechanical energy of the VFD could be misdirected to the other >96% of proteins present.

The paradigm introduced here, applying shear force to drive rapid equilibration of protein folding has not been explored in biochemistry, yet is extensively applied in the cell by chaperonins. Another principle of chaperonin function, avoiding reaggregation by cage-like isolation during refolding, also applies to the VFD, which relies upon shear forces to maintain protein separation[18,33] .

3.6 Discussion

This report offers significant advantages over conventional approaches to protein refolding. First, VFD-mediated refolding requires much smaller solution volumes (approximately 1% of the volumes required for conventional dialysis). Second, this key step in protein production occurs >100-times faster than overnight dialysis with >1000-fold improvements for proteins such as a caveolin. Notably, introducing high shear in thin fluid films is a low energy, inexpensive process.

The advantages of VFD refolding open new possibilities for increasing protein yields in simple cell lines. The VFD can untangle complex mixtures, aggregates and insoluble inclusion bodies. Illustrating this advantage, high concentrations of a chemical inducer like IPTG could drive overexpressed proteins into insoluble inclusion bodies.

Most processes would avoid inclusion bodies by optimization of growth conditions and special cell lines at the expense of higher yields and purer protein. Furthermore, the continuous flow mode of the VFD readily allows scale-up to accommodate much larger solution volumes, and the approach could drastically lower the time and financial costs required to refold inactive proteins at an industrial scale. The VFD sample tube itself can also be modified to amplify or otherwise direct the intensity of shear forces applied; for example, modified surfaces with high contact angle and/or with textured features could enhance the turbulent flow, altering the applied shear stress. Harnessing shear forces to achieve rapid equilibration of protein folding could be expanded to a wide-range of applications for research and manufacturing.

3.7 Materials and Methods

All buffers are listed in Tables 3-1 and 3-2.

Protein	*Expression time (h)*	*Expression temp. (°C)*	*[IPTG] (M)*	*Lysis buffer*
Hen egg white lysozyme	4	37	1	50 mM NaH_2PO_4, 500 mM NaCl, 10 mM imidazole, 1 mM HALT protease inhibitor (Pierce), 10 mM 2-mercaptoethanol, pH 8.0
Caveolin-ΔTM	3	37	0.5	50 mM Tris-HCl, 10 mM NaCl, 5 mM EDTA, 100 mM PMSF, pH 8.0
HIV gp41	8	22	0.5	50 mM NaH_2PO_4, 300 mM NaCl, 10 mM 2-mercaptoethanol, 1 mM HALT protease inhibitor, pH 8.0
His-PKA	5	37	1	50 mM NaH_2PO_4, 500 mM NaCl, 10 mM imidazole, 1 mM HALT protease inhibitor, 10 mM 2-mercaptoethanol, pH 8.0

Table 3-1. Expression conditions of recombinantly expressed proteins.

Protein	*Resin*	*Denaturing buffer*	*Binding/wash buffer*	*Elution buffer*
HEWL	UNOsphere S (Bio-Rad)	20 mM Tris, 10 mM NaCl, 8 M urea, pH 7.8	20 mM Tris, 10 mM NaCl, 8M urea, pH 7.8	Wash buffer, 400 mM NaCl
Caveolin-ΔTM	Ni-NTA (Bio-Rad)	50 mM Tris, 50 mM NaCl, 8 M urea, pH 8.0	50 mM Tris-HCl, 300 mM NaCl, 10 mM imidazole, 0.2% Na azide, 8M urea, pH 8.0	Wash buffer, pH 4.0
HIV gp41	Ni-NTA	-	50 mM NaH_2PO_4, 300 mM NaCl, 10 mM 2-mercaptoethanol, 20 mM imdazole, pH 8.0	Wash buffer, 250 mM imidazole
His-PKA	Ni-NTA	20 mM NaH_2PO_4, 500 mM NaCl, 6 M guanidine-HCl, pH 7.0	50 mM NaH_2PO_4, 300 mM NaCl, 1 mM imdazole, pH 7.0	Wash buffer, 500 mM imidazole

Table 3-2. Purification conditions of recombinantly expressed proteins. The protein target used for ELISA binding studies, HIV gp41, was purified with non-denaturing conditions.

3.7.1 Expression and purification of HEWL, caveolin-ΔTM, and HIV gp41

Hen egg white lysozyme (HEWL), caveolin-ΔTM, and HIV gp41 were overexpressed in BL21 *E. coli* by induction with isopropyl β-D-1-thiogalactopyranoside (IPTG, 1 mM). The 1 l culture volume was centrifuged at 6000 rpm to collect the bacterial pellet. The pellet was reconstituted in lysis buffer and sonicated in 30 s continuous bursts with 1 min cooling on ice for eight cycles (20 watts). HEWL and caveolin-ΔTM, were purified under denaturing conditions, and HIV gp41 was purified under non-denaturing conditions. For specific expression and purification conditions, see Table 3-1 and 3-2, respectively. The egg whites were obtained from chicken eggs, and diluted 2:3 in PBS, heat-treated at 90 °C for 20 min, and dissolved in 8 M urea overnight at 4 °C. The His_6 tag was cleaved from HIV gp41 with Tobacco Etch Virus protease, which was then removed by IMAC. All protein concentrations were determined by bicinchoninic acid assay kit (Pierce).

3.7.2 Protein refolding and determination of native state confirmation

Commercial, lyophilized HEWL protein (Sigma) was reconstituted in PBS as 'active' HEWL sample. Recombinantly expressed HEWL was pre-treated by 1:100 rapid dilution in PBS, and then refolded by VFD treatment. All samples were treated at 22 °C within a 16 cm long, 10 mm diameter glass test tube. When operated in confined mode, the VFD was set to a 45° tilt angle and 1 ml was spun at 5 krpm, unless otherwise noted. The continuous mode experiment was conducted by flowing the rapidly diluted protein through the inlet port to the base of the sample at a flow rate of 0.1 ml/min. Caveolin-ΔTM VFD refolding was performed in confined mode (1 ml, 5 krpm, 22 °C).

For comparison, caveolin-ΔTM was also refolded using conventional dialysis over 4 days (1:500, 50 mM Tris-HCl, 1 mM EDTA, 4 °C, pH 8.5).

Circular dichroism spectra of HEWL were collected immediately following VFD refolding in PBS (20 nm/min, 4 accumulations), and caveolin-ΔTM were collected in 10 mM sodium phosphate, pH 7.5 (10 nm/min, 8 accumulations). All lysozyme activity assays used the EnzChek kit (Invitrogen) after rapid dilution from denatured protein solution into PBS (1:100) according to manufacture instructions, except for decreasing the 37 °C incubation time from 30 to 10 min. Lysozyme activity was interpolated by least-square regressions fit of lysozyme standards to a Michaelis-Menten curve ($v = \frac{V_{max}[S]}{K_m + [S]} + c$) with Prism 6 software (GraphPad, Figure 3-2). After 1:100 rapid dilution into PBS, protein solution contains 80 mM urea, 0.2 mM Tris and 4 mM NaCl. Lysozyme activity with various levels of urea, Tris, and NaCl was determined to verify that the assay was not affected (Figure 3-6). Purified caveolin-ΔTM from the inclusion body were diluted 1:5 in 10 mM sodium phosphate, pH 7.5 and then briefly dialyzed prior to VFD treatment (1:100 volume, 1 h, 4 °C) for circular dichroism spectra.

3.7.3 ELISA binding assays

The dose-dependent ELISA was conducted by coating HIV gp41 (100 µl of 10 µg/ml in 50 mM sodium carbonate pH 9.6 for 4 h at 4 °C) on a Nunc Maxisorp 96-well microtiter plate. After removing the coating solution, a blocking solution of 0.2% non-fat milk in PBS was applied. Caveolin-ΔTM, anti-His mouse monoclonal antibody (Sigma, H1029), and anti-mouse HRP-conjugated polyclonal antibody (1:2000, Sigma, A5906) were diluted in 100 µl PT buffer (1:1000, PBS, 0.05% Tween-20) and incubated for 1 h

at 4 °C with four wash steps using PT buffer (200 µl). The ELISA was developed by the addition of 1% w/v o-phenylenediamine dihydrocholoride in citric acid buffer (0.02% w/v H_2O_2, 50 mM citric acid, 50 mM Na_2HPO_4, pH 5.0), and the absorbance of the solution was measured at 450 nm using a microtiter plate reader.

3.7.4 Shear stress-mediated refolding of His-PKA bound IMAC resin

The catalytic subunit of PKA was overexpressed in BL21 *E. coli* with an *N*-terminal His_6 tag by induction with IPTG (1 mM). This experiment applied the residual pellet from a 12 l culture, a waste product more typically discarded. After dissolution in lysis buffer, sonication was applied as described above. His-PKA was then denatured in 6 M guanidine-HCl, 20 mM sodium phosphate, 500 mM NaCl and incubated with Ni^{2+}-charged Profinity IMAC resin (Bio-Rad) for 2 h at room temperature (1 ml of a 1.72 mg/ml His-PKA to 50 µl or 250 µl bed volume IMAC). A control experiment used uncharged IMAC resin instead. The IMAC-His-PKA solution was then diluted to 1 M guanidine-HCl with binding buffer containing 1 mM imidazole, or with the elution buffer containing 500 mM imidazole as a control. This diluted solution was immediately treated in the VFD (1 ml, 5 krpm, 20 min). After transferring to a 1.5 ml eppendorf tube, the resin was washed by aliquoting 1 ml wash buffer, inverting the tube three times, and centrifuging the tube at 2000 $\times$ g for 2 min to separate the beads from the supernatant. This process was repeated two additional times before elution with elution buffer containing 500 mM imidazole. For protein quantification only, samples containing 500 mM imidazole were diluted 1:100 in wash buffer to prevent residual imidazole from interfering with the BCA assay. PKA activity was determined by monitoring substrate

depletion in an NADH enzyme-linked assay at 340 nm (300 µl assay volume, 10 mM ATP, 0.5 mM NADH, 1 mM phosphoenolpyruvate, 0.0153 U/µl lactate dehydrogenase, 0.0269 U/µl pyruvate kinase, 0.67 mM kemptide, 100 mM MOPS, 9 mM $MgCl_2$, pH 7.0). Kemptide was synthesized by solid-phase peptide synthesis. All other reagents were purchased from Sigma-Aldrich.

REFERENCES

1. Meyer, H. & Schmidhalter, D. in *Innov. Biotechnol.* 211–250 (2012).

2. Freire, E., Schön, A. & Velazquez-Campoy, A. Isothermal titration calorimetry: general formalism using binding polynomials. *Methods Enzymol.* **455,** 127–55 (2009).

3. Futaki, S. Membrane-permeable arginine-rich peptides and the translocation mechanisms. *Adv. Drug Deliv. Rev.* **57,** 547–558 (2005).

4. Radford, S. E. Protein folding: progress made and promises ahead. *Trends Biochem. Sci.* **25,** 611–618 (2000).

5. Jemth, P. *et al.* Demonstration of a low-energy on-pathway intermediate in a fast-folding protein by kinetics, protein engineering, and simulation. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 6450–6455 (2004).

6. Wang, L., Maji, S. K., Sawaya, M. R., Eisenberg, D. & Riek, R. Bacterial inclusion bodies contain amyloid-like structure. *PLoS Biol.* **6,** e195 (2008).

7. Wang, L., Schubert, D., Sawaya, M. R., Eisenberg, D. & Riek, R. Multidimensional structure-activity relationship of a protein in its aggregated states. *Angew. Chem. Int. Ed. Engl.* **49,** 3904–3908 (2010).

8. Vallejo, L. F. & Rinas, U. Strategies for the recovery of active proteins through refolding of bacterial inclusion body proteins. *Microb. Cell Fact.* **3,** 11 (2004).

9. Tsumoto, K., Ejima, D., Kumagai, I. & Arakawa, T. Practical considerations in refolding proteins from inclusion bodies. *Protein Expr. Purif.* **28,** 1–8 (2003).

10. Li, F., Vijayasankaran, N., Shen, A., Kiss, R. & Amanullah, A. Cell culture processes for monoclonal antibody production. *MAbs.* **2,** 466–479 (2010).

11. Hannig, G. & Makrides, S. C. Strategies for optimizing heterologous protein expression in Escherichia coli. *Trends Biotechnol.* **16,** 54–60 (1998).

12. Andersen, D. C. & Krummen, L. Recombinant protein expression for therapeutic applications. *Curr. Opin. Biotechnol.* **13,** 117–123 (2002).

13. St John, R. J., Carpenter, J. F. & Randolph, T. W. High pressure fosters protein refolding from aggregates at high concentrations. *Proc. Natl. Acad. Sci. U. S. A.* **96,** 13029–13033 (1999).

14. Qoronfleh, M. W., Hesterberg, L. K. & Seefeldt, M. B. Confronting high-throughput protein refolding using high pressure and solution screens. *Protein Expr. Purif.* **55,** 209–24 (2007).

15. Mayhew, M. & Silva, A. da. Protein folding in the central cavity of the GroEL–GroES chaperonin complex. *Nature.* **379,** 420–426 (1996).

16. Walter, S. & Buchner, J. Molecular chaperones--cellular machines for protein folding. *Angew. Chemie Int. Ed.* **41,** 1098–1113 (2002).

17. Hartl, F. U. & Hayer-Hartl, M. Converging concepts of protein folding in vitro and in vivo. *Nat. Struct. Mol. Biol.* **16,** 574–581 (2009).

18. Hartl, F. U., Bracher, A. & Hayer-Hartl, M. Molecular chaperones in protein folding and proteostasis. *Nature.* **475,** 324–332 (2011).

19. Lin, Z., Madan, D. & Rye, H. S. GroEL stimulates protein folding through forced unfolding. *Nat. Struct. Mol. Biol.* **15,** 303–311 (2008).

20. Chakraborty, K. *et al.* Chaperonin-catalyzed rescue of kinetically trapped states in protein folding. *Cell.* **142,** 112–122 (2010).

21. Brockwell, D. J. & Radford, S. E. Intermediates: ubiquitous species on folding energy landscapes? *Curr. Opin. Struct. Biol.* **17,** 30–37 (2007).

22. Hartl, F. Molecular chaperones in cellular protein folding. *Nature* **381,** 571–579 (1996).

23. Chen, X., Dobson, J. F. & Raston, C. L. Vortex fluidic exfoliation of graphite and boron nitride. *Chem. Commun.* **48,** 3703–3705 (2012).

24. Wahid, M. H., Eroglu, E., Chen, X., Smith, S. S. M. & Raston, C. C. L. Functional multi-layer graphene–algae hybrid material formed using vortex fluidics. *Green Chem.* **15,** 650–655 (2013).

25. Boulos, R. A. *et al.* Spinning up the polymorphs of calcium carbonate. *Sci. Rep.* **4,** 3616 (2014).

26. Mallock, A. Determination of the viscosity of water. *Proc. R. Soc. A.* **45,** 126–132 (1888).

27. Rayleigh, L. On the Dynamics of Revolving Fluids. *Proc. R. Soc. A.* **93,** 148–154 (1917).

28. Bekard, I. B., Asimakis, P., Bertolini, J. & Dunstan, D. E. The effects of shear flow on protein structure and function. *Biopolymers.* **95,** 733–745 (2011).

29. Saxena, V. P. & Wetlaufer, D. B. A new basis for interpreting the circular dichroic spectra of proteins. *Proc. Natl. Acad. Sci.* **68,** 969–972 (1971).

30. Hovanessian, A. G. *et al.* The caveolin-1 binding domain of HIV-1 glycoprotein gp41 is an efficient B cell epitope vaccine candidate against virus infection. *Immunity* **21,** 617–627 (2004).

31. Wang, X. M., Nadeau, P. E., Lo, Y.-T. & Mergia, A. Caveolin-1 modulates HIV-1 envelope-induced bystander apoptosis through gp41. *J. Virol.* **84,** 6515–6526 (2010).

32. Evans, M. S., Sander, I. M. & Clark, P. L. Cotranslational folding promotes beta-helix formation and avoids aggregation in vivo. *J. Mol. Biol.* **383,** 683–692 (2008).

33. White, D. A., Buell, A. K., Knowles, T. P. J., Welland, M. E. & Dobson, C. M. Protein aggregation in crowded environments. *J. Am. Chem. Soc.* **132,** 5170–5175 (2010).

CHAPTER 4

Studies towards directed evolution of alternate non-immunoglobulin scaffolds with filamentous M13 phage display

Abstract

Exploring new protein scaffolds expands the repertoire of protein libraries for use in diagnostic and therapeutic functions. Specifically, non-immunoglobulin based scaffolds can avoid complications associated with traditional antibody development, including, but not limited to, expression yields, proper glycosylation, and intellectual property considerations. Several classes of these alternative protein scaffolds have proven to be commercially viable, with affinities and specificities comparable to antibodies. We report the adaptation of echistatin, a 49-residue disintegrin protein, as a new protein ligand scaffold. The exceptionally small size of echistatin makes the protein an ideal target for M13 filamentous phage display, chemical synthesis, and recombinant expression. Selections with the phage-displayed echistatin library successfully identified ligands to the cancer-associated biomarker prostate specific antigen.

Introduction

Prostate cancers remain the most commonly diagnosed carcinomas in men. In 2010, more than 28,000 men in the United States succumbed to the disease, making it the deadliest cancer among American men[1]. The U.S. Food and Drug Administration approved screenings for prostate specific antigen (PSA) to facilitate early diagnosis of prostate cancer[2]. Malignant prostate cells excrete highly elevated concentrations of this 32 kDa serine protease[3,4]. However, recent guidance issued by the United States Preventive Services Task Force recommended halting regular PSA screenings to address the overdiagnosis and overtreatment of slow-growing asymptomatic tumors[2]. Despite this, PSA remains an important biomarker to track recurrent cases of prostate cancer. Thus, progress towards the development of a cheap, easily expressed protein ligand to PSA can drive down the costs associated with regular prostate cancer screening.

Existing federally approved tests for prostate specific antigen rely on the production of anti-PSA monoclonal antibodies[5,6]. Unfortunately, immunoglobulins are comparatively expensive to manufacture and are burdened by a complex patent landscape. An alternative approach utilizes modified non-immunoglobulin scaffolds to bind to the cancer biomarker of interest. Notably, protein ligands derived from ubiquitin[7], protein A derived affibodies[8], and designed ankryin repeat proteins[9] have been successfully commercialized. These small, low molecular weight proteins constitute a family of antibody mimetics that have been engineered to bind to a wide variety of target proteins, but do not share structural similarities to their immunoglobulin counterparts[10,11]. Unlike antibodies, these alternative protein scaffolds can be

expressed in large quantities in bacterial hosts. In this study, we report the design and selection of a small disintegrin protein echistatin[12] library on M13 bacteriophage to identify a high-affinity ligand to the cancer-associated biomarker PSA. The ligands identified here may contribute to the development of cheap, readily available diagnostics to address the ongoing need for recurrent prostate cancer detection.

4.1 Echistatin library design and selection

Echistatin is a small disintegrin protein from the venomous viper *Echis carinatus.* While echistatin has been adapted for use as an anti-angiogenesis therapeutic to delay tumor growth and increase survival[13,14], it remains unexplored as a protein ligand scaffold despite its exceptionally small size and low complexity. Unlike most protein biopharmaceuticals, echistatin is very small and can be chemically synthesized[15].

To facilitate the expression and selection of a large echistatin library, we utilized M13 filamentous phage display[16]. The gene encoding for the target of choice is subcloned into the phage plasmid, and the resultant protein expressed as fusions to the P3 or P8 coat proteins of the phage[17,18]. The Kunkel method[19] of site-directed mutagenesis was utilized to generate a library of echistatin variants. Two loops near the C-terminus of the protein were chosen for diversification into a mixture of the 20 naturally occurring amino acids. The first loop is comprised of a 9-residue span, and the second loop replaced five adjacent amino acids. In wild-type echistatin both regions primarily consist of hydrophilic residues, as expected for solvent-exposed loops. Following diversification, the phage-displayed library was subjected to several rounds of selection and propagation to enrich for high-affinity binders to PSA[18]. Sequencing the

phage plasmid enables rapid identification of the selected variants[20]. An overview of the phage display selection is shown in Figure 4-1.

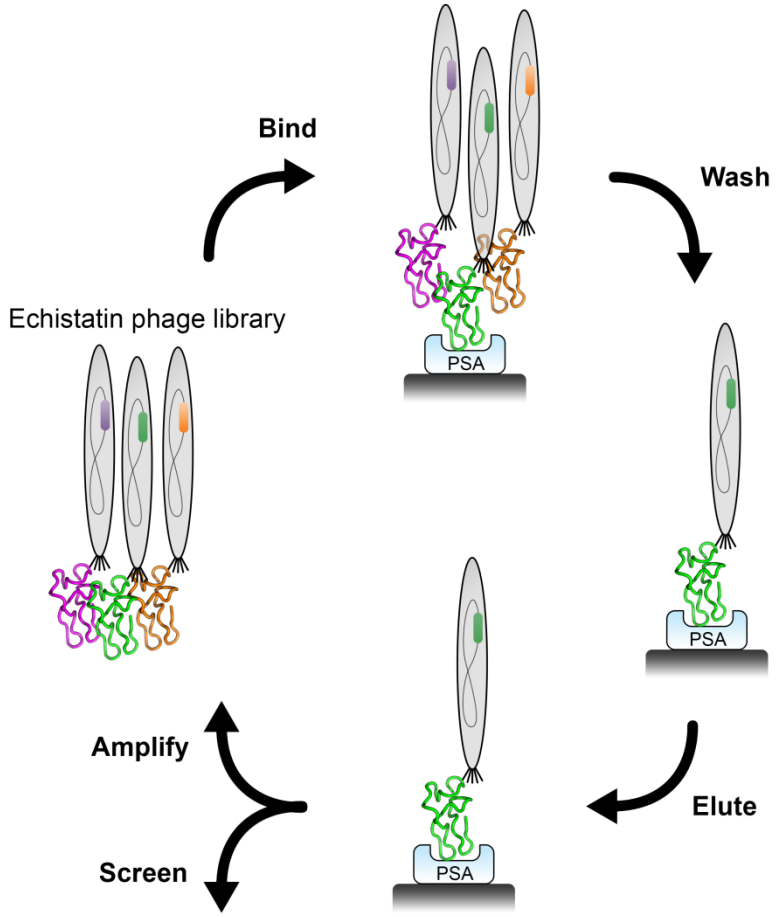


Figure 4-1. General schematic of selections with phage-displayed echistatin proteins. The echistatin gene is inserted into the phage genome as part of the open reading frame encoding the P3 coat protein. Each echistatin variant is displayed as a fusion to one of the five P3 proteins during phage propagation. Once a library of phage-displayed echistatin variants is established, four rounds of selections and amplifications were performed to identify binders to the immobilized prostate specific antigen target. Weak binders are removed during the wash steps. Following the last round of selections, variants are screened in ELISA binding assays.

4.2 Results and Discussion

Selections were conducted by immobilizing PSA in 96-well microtiter plates and panning for binders by an affinity-based methodology[21]. The phage-displayed echistatin variants compete for a limited amount of PSA and the highest affinity variant outcompetes its less optimized siblings. Four rounds of selection and amplification were conducted. PCR amplification and sequencing of selected echistatin variants identified two candidates with improved binding characteristics to PSA (Figure 4-2). Variant 10 echistatin contained a single amino acid substitution to glutamate. In contrast, variant 39 echistatin replaces the predominantly hydrophilic loop 1 with hydrophobic leucines, tyrosines, and tryptophans. Similar amino acid compositions are well represented in protein-protein binding interfaces such as antibody-antigen binding contacts[22], engineered adnectins ligands[23], and diversified Mtd libraries[24]. Additionally, variant 39 contains a mutation at cysteine-32 that disrupts one of the four disulfide bonds. This position was not part of the original library design but may have been introduced during PCR amplification of the echistatin gene by *Taq* polymerase, which lacks 3' to 5' exonuclease activity[25], or mutagenesis during amplification in *E. coli*.

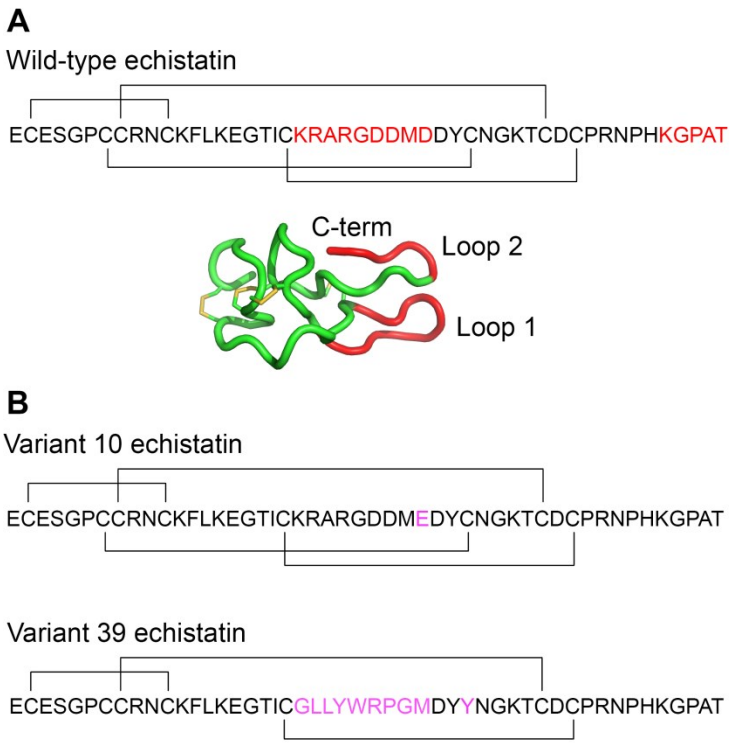


Figure 4-2. Echistatin library design and selection. Amino acid sequences of (A) wild-type echistatin and (B) variants identified from selections for binding to prostate specific antigen. Disulfide bonds are indicated by the connecting lines above and below the sequence. Residues selected for diversification are highlighted in red. A ribbon representation of echistatin is shown with mutagenized loops 1 and 2 also in red (PDB: 1RO3)[26]. Deviations from the wild-type sequence in variant 10 and variant 39 echistatins are highlighted in magenta.

Variant 10 and variant 39 echistatins do not share any common amino acid substitutions but curiously demonstrate similar levels of binding to PSA in preliminary ELISAs (Figure 4-3A). In competitive ELISA experiments, the phage-displayed variant 10 echistatin bound to PSA with an apparent K_D of 1.8 nM while the phage-displayed variant 39 echistatin bound with an apparent K_D of 0.1 nM (Figure 4-3B). The low apparent dissociation constants calculated are likely a product of high avidity[27] mediated by multivalent binding of the phage-displayed echistatin to the surface of immobilized PSA. Studies utilizing isothermal titration calorimetry will need to be completed to

resolve the dissociation constant between the echistatin macromolecules to PSA in solution.

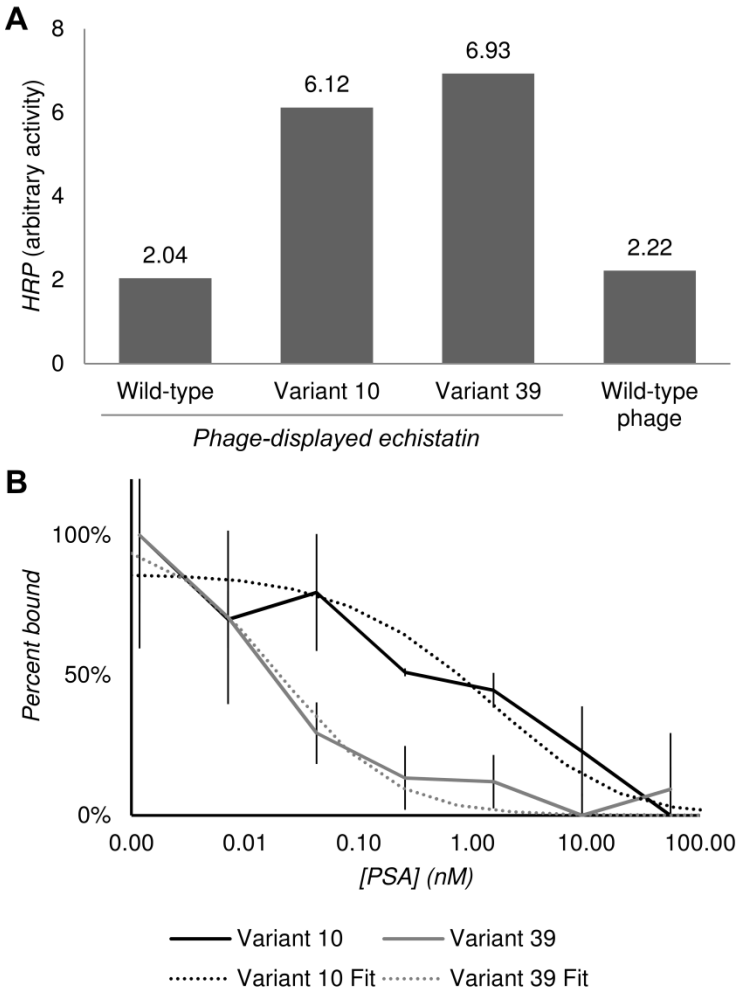


Figure 4-3. Phage-displayed echistatin binding assays. (A) ELISA experiments demonstrate that phage-displayed variants 10 and 39 bind with higher affinity to prostate specific antigen (coated at 76 nM) relative to phage-displayed wild-type echistatin and wild-type M13 phage. (B) Competitive ELISAs demonstrate that the two phage-displayed echistatin variants bind with high-affinity to the PSA antigen. Apparent dissociation constants were calculated at 1.8 nM and 0.1 nM for phage-displayed variant 10 and variant 39 echistatins, respectively.

4.3 Future directions

The binding studies described above suggest echistatin can be added to the repertoire of non-immunoglobulin protein scaffolds. However, selections for other protein targets should also be performed to fully validate this small disintegrin protein as a general protein-binding ligand. Affinity maturation of the two echistatin variants can further improve the binding characteristics to PSA[21]. Furthermore, homolog shotgun scanning, a technique pioneered in the Weiss lab[28], could map the side-chain contributions to binding.

Fortunately, we can further extend the functionality and sensitivity of the echistatin variants already identified by utilizing methods developed in the Weiss lab. Mohan and colleagues recently demonstrated increased sensitivity to the cancer biomarker prostate specific membrane antigen by bivalent binding of two distinct peptide ligands, attached to the same phage particle, to prostate specific membrane antigen (PSMA)[29]. In this technique, the first ligand is genetically encoded in the phage genome and is expressed as fusions to the phage coat proteins, similar to traditional phage display. The second ligand associates non-covalently to the negatively charged P8 coat protein following conjugation to a 14-lysine repeat. This same technique could be used to increase binding affinity to prostate specific antigen by expressing variant 10 echistatin as a fusion to the P3 coat protein followed by attachment of a polylysine - variant 39 echistatin fusion to the P8 coat protein.

Furthermore, researchers in Weiss lab, in collaboration with the Penner group at the University of California, Irvine, have integrated whole-phage ligands in conducting polymer films to detect antigen in real-time[29–32]. Incorporation of the whole virus particle

displaying a peptide ligand to PSMA enabled detection of the antigen in the picomolar range[29]. Further studies are required to establish if the phage-polymer films can tolerate the larger echistatin protein in place of the peptide ligands used in PSMA diagnostic devices.

As additional novel techniques are explored to facilitate the generation of high-affinity binders, well-established scaffold systems will continue to play an important role in the development of therapeutic and diagnostic biopharmaceuticals. The addition of new scaffolds will enable investigators to improve binding characteristics and expand the gamut of protein, peptide, nucleic acid and small molecule targets accessible by directed evolution methods[23].

4.4 Materials and Methods

4.4.1 Construction of the phage-displayed echistatin libraries

Design and construction of the phage-displayed echistatin libraries was completed by Dr. Cathie Overstreet. The wild-type cDNA encoding echistatin was PCR amplified with NsiI and FseI restriction sites along with an N-terminal FLAG epitope. Following double digestion of the PCR product and the phage display vector, the digested products were ligated with T4 DNA ligase. Next, the ligated product was transformed into calcium competent *E. coli* XL-1 cells and selected for proper uptake of the ligated phagemid by plating on LB plates with carbenicillin (50 μg/ml). The phagemid was then purified with a plasmid miniprep kit (Zymo).

The phage-displayed library of echistatin variants was produced using the Kunkel method[33] of site-directed mutagenesis with NNS degenerate codons corresponding to

residues 21 – 29 and 45 – 49 (Table 4-1). Following previously published methods[18], the phage library was propagated with KO7 helper phage in 4 x 70 ml cultures (2YT, 50 µg/ml carbenicillin, 20 µg/ml kanamycin, 37° C, 16 h). Phage particles were isolated with PEG-NaCl precipitation and resuspended in PBS.

Degenerate Oligonucleotide Table

Target Residues	*Sequence*	*Restriction sites*
21 - 29	GAA GGA ACA ATT TGT NNS NNS NNS NNS NNS NNS NNS NNS NNS GAT TAT TGT AAT GGA	NsiI / FseI
45 - 49	CCT CGG AAC CCC NNS NNS NNS NNS NNS NNS GGC CGG CCC TC	NsiI / FseI

Table 4-1. Degenerate oligonucleotides utilized for echistatin library construction. The NNS codon allows substitution of all 20 amino acids and also encodes for the TAG stop codon.

4.4.2 Selections with phage-displayed echistatin

Maxisorp 96-well microtiter plates (Nunc) were coated with 200 µl of purified prostate specific antigen (76 nM, 50 mM $NaCHO_3$, pH 9.6, 1 h) and blocked with 0.2% non-fat milk in PBS. Following removal of the blocking agent, a 1:1 mixture of the purified phage and blocking agent were added to each well (200 µl, 1 h). The wells were then washed with PBS + 0.05% Tween-20 (300 µl). The wash counts were 3, 6, 9, and 12 for each of the four selection rounds, respectively. Phage that remained bound after the wash steps were eluted with 0.1 M HCl on a sonication water bath (100 µl, 10 min). Each well was then neutralized with the addition of 1 M Tris-HCl (33 µl, pH 8). Biopanned phage variants were then pooled and re-propagated in *E. coli* XL-1 cells with KO7 helper phage. Following the 4^{th} round of selections, phage titers were used for colony PCR and echistatin genes were sequenced (Genewiz).

4.4.3 ELISA binding studies

The two phage-displayed echistatin mutants variant 10 and variant 39 were propagated in XL-1 and purified by PEG-NaCl precipitation as described above. Maxisorp microtiter plates were coated with PSA and blocked in the same fashion as the selections. The purified phage was then added to the wells and allowed to incubate for 1 h (16 nM, PBS, 0.2% non-fat milk, 100 µl). Each well was washed 3 times with PBS + 0.05% Tween-20 (300 µl). A horseradish peroxidase-conjugated anti-M13 antibody (1:2500, PBS, GE Healthcare) detected bound phage particles. The plate was developed with 1% w/v *o*-phenylenediamine dihydrocholoride in citric acid buffer (0.02% w/v H_2O_2, 50 mM citric acid, 50 mM Na_2HPO_4, pH 5.0) and HRP activity was determined by absorbance measurements at 450 nm.

Competitive ELISA studies also utilized PSA-coated Maxisorp plates but the phage-displayed echistatin variants were pre-incubated with the indicated concentrations of PSA for 2 hours (PBS, 0.2% non-fat milk, 100 µl). After the plate is blocked and the blocking agent is removed, the phage-PSA solution was aliquoted into the microtiter plate and allowed to incubate for 1 h. The remaining levels of bound phage are detected with the same steps as described above. HRP activity measurements were normalized and dissociation constants were calculated from least-squares fit as described in previously published methods[34].

REFERENCES

1. Department of Health and Human Services, Centers for Disease Control and Prevention & National Cancer Institute. United States Cancer Statistics: 1999–2010 Incidence and Mortality Web-based Report. *http://www.cdc.gov/uscs* (2013).

2. US Preventive Services Task Force. *Talking With Your Patients About Screening for Prostate Cancer*. 1–2 (2012).

3. Catalona, W. *et al.* Comparison of digital rectal examination and serum prostate specific antigen in the early detection of prostate cancer: results of a multicenter clinical trial of 6,630 men. *J. Urol.* **151,** 1283–1290 (1994).

4. Lilja, H. Biology of prostate-specific antigen. *Urology.* **62,** 27–33 (2003).

5. Dnistrian, A. M., Schwartz, M. K., Smith, C. A., Nisselbaum, J. S. & Fair, W. R. Abbott IMx evaluated for assay of prostate-specific antigen in serum. *Clin. Chem.* **38,** 2140–2142 (1992).

6. Laffin, R. J. *et al.* Hybritech Total and Free Prostate-specific Antigen Assays Developed for the Beckman Coulter Access Automated Chemiluminescent Immunoassay System: A Multicenter Evaluation of Analytical Performance. *Clin. Chem.* **47,** 129–132 (2001).

7. Lorey, S. *et al.* Novel ubiquitin-derived high affinity binding proteins with tumor targeting properties. *J. Biol. Chem.* **289,** 8493–8507 (2014).

8. Wikman, M. *et al.* Selection and characterization of HER2/neu-binding affibody ligands. *Protein Eng. Des. Sel.* **17,** 455–462 (2004).

9. Stumpp, M. T., Binz, H. K. & Amstutz, P. DARPins: a new generation of protein therapeutics. *Drug Discov. Today.* **13,** 695–701 (2008).

10. Skerra, A. Alternative non-antibody scaffolds for molecular recognition. *Curr. Opin. Biotechnol.* **18,** 295–304 (2007).

11. Binz, H. K., Amstutz, P., Plückthun, A. & Pluckthun, A. Engineering novel binding proteins from nonimmunoglobulin domains. *Nat. Biotechnol.* **23,** 1257–1268 (2005).

12. Gan, Z. R., Gould, R. J., Jacobs, J. W., Friedman, P. A. & Polokoff, M. A. Echistatin. A potent platelet aggregation inhibitor from the venom of the viper, Echis carinatus. *J. Biol. Chem.* **263,** 19827–19832 (1988).

13. Santulli, G. *et al.* Evaluation of the anti-angiogenic properties of the new selective αVβ3 integrin antagonist RGDechiHCit. *J. Transl. Med.* **9,** 7 (2011).

14. Minea, R. *et al.* Development of a chimeric recombinant disintegrin as a cost-effective anti-cancer agent with promising translational potential. *Toxicon.* **59,** 472–486 (2012).

15. Garsky, V. M. *et al.* Chemical synthesis of echistatin, a potent inhibitor of platelet aggregation from Echis carinatus: synthesis and biological activity of selected analogs. *Proc. Natl. Acad. Sci. U. S. A.* **86,** 4022–4026 (1989).

16. Smith, G. P. & Petrenko, V. A. Phage Display. *Chem. Rev.* **97,** 391–410 (1997).

17. Smith, G. P. & Petrenko, V. a. Phage Display. *Chem. Rev.* **97,** 391–410 (1997).

18. Sidhu Weiss, Gregory A., S. S. in *Phage Disp.* (Clackson Lowman, Henry B., T.) **266,** 27–41 (Oxford University Press, 2004).

19. Kunkel, T. A. Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proc. Natl. Acad. Sci. U. S. A.* **82,** 488–492 (1985).

20. Smith, G. P. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science.* **228,** 1315–1317 (1985).

21. Levin, a M. & Weiss, G. a. Optimizing the affinity and specificity of proteins with molecular display. *Mol. Biosyst.* **2,** 49–57 (2006).

22. Koide, S. & Sidhu, S. S. The Importance of Being Tyrosine: Lessons in Molecular Recognition from Minimalist Synthetic Binding Proteins. *ACS Chem. Biol.* **4,** 325–334 (2009).

23. Gilbreth, R. N. & Koide, S. Structural Insights for Engineering Binding Proteins Based on Non-Antibody Scaffolds. *Curr. Opin. Struct. Biol.* **22,** 413–420 (2013).

24. Yuan, T. Z., Overstreet, C. M., Moody, I. S. & Weiss, G. A. Protein Engineering with Biosynthesized Libraries from Bordetella bronchiseptica Bacteriophage. *PLoS One.* **8,** e55617 (2013).

25. Lawyer, F. C. *et al.* High-level expression, purification, and enzymatic characterization of full-length Thermus aquaticus DNA polymerase and a truncated form deficient in 5’ to 3' exonuclease activity. *PCR Methods Appl.* **2,** 275–287 (1993).

26. Esteve, V., Kovacs, H., Calvete, J. J., Celda, B. & Monleón, D. Conformation and concerted dynamics of the integrin-binding site and the C-terminal region of echistatin revealed by homonuclear NMR. *Biochem. J.* **387,** 57–66 (2005).

27. Krishnamurthy, V. M., Estroff, L. A. & Whitesides, G. M. in *Fragm. Approaches Drug Discov.* 11–52 (2006).

28. Levin, A. M. *et al.* Double barrel shotgun scanning of the caveolin-1 scaffolding domain. *ACS Chem. Biol.* **2,** 493–500 (2007).

29. Mohan, K., Donavan, K. C., Arter, J. A., Penner, R. M. & Weiss, G. A. Sub-nanomolar Detection of Prostate-Specific Membrane Antigen in Synthetic Urine by Synergistic, Dual-Ligand Phage. *J. Am. Chem. Soc.* **135,** 7761–7767 (2013).

30. Weiss, G. A. & Penner, R. M. The promise of phage display: customized affinity and specificity. *Anal. Chem.* **80,** 933–943 (2008).

31. Arter, J. A., Taggart, D. K., McIntire, T. M., Penner, R. M. & Weiss, G. A. Virus-PEDOT nanowires for biosensing. *Nano. Lett.* (2010). doi:10.1021/nl1025826

32. Arter, J. a *et al.* Virus-polymer hybrid nanowires tailored to detect prostate-specific membrane antigen. *Anal. Chem.* **84,** 2776–83 (2012).

33. Kunkel, T. A., Roberts, J. D. & Zakour, R. A. Rapid and efficient site-specific mutagenesis without phenotypic selection. *Methods Enzym.* **154,** 367–382 (1987).

34. Martineau, P. *Affinity Measurements by Competition ELISA*. *Antib. Eng.* **1,** (Springer, 2010).

CHAPTER 5

Conclusions and future directions

While protein engineering methods utilizing *Bordetella* phage display libraries have shown some success, several shortfalls remain unaddressed. In *Bordetella* phage display, repeat selections are counter-productive due to the self-mutating nature of the host phage. If the diversity generation could be dynamically fine-tuned, or temporarily halted, experimenters could alternate rounds of selection and enrichment with rounds of mutagenesis. This switching methodology more closely mimics the affinity maturation process applied to vertebrate immunoglobulins (Figure 5-1). This may be possible by reversibly inhibiting the phage-encoded reverse transcriptase.

Another possibility is to disrupt the mutagenic retrohoming required to guide the diversified cDNA into the coding variable region of *mtd*. Alayyoubi and colleagues identified an essential protein, *Bordetella* accessory variability determinant (bAvd), that appears to function as a nucleic acid chaperone to help mediate this process and is an interesting target for inhibition studies[1,2].

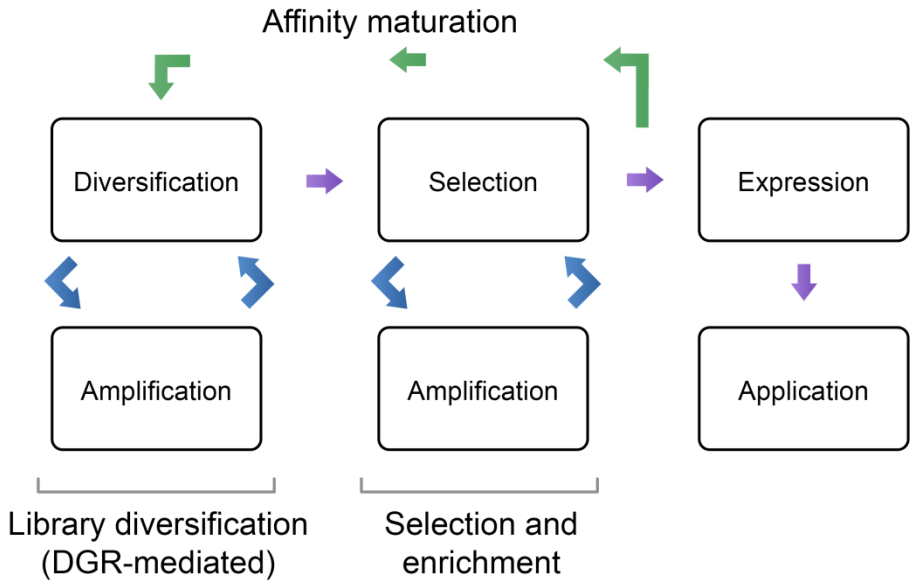


Figure 5-1. *Bordetella* phage display strategy with dynamic DGR switching. The protein scaffold undergoes multiple DGR-mediated diversification and amplification rounds before the first selection step. Following diversification, the DGR is temporarily inhibited to allow selection and amplification rounds to enrich the protein library for ligands to a predetermined target. Following enrichment, the library can then undergo affinity maturation through DGR-mediated mutagenesis.

Furthermore, Guo *et al*. discovered GC-rich inverted repeats downstream of the *vr* gene form hairpin structures critical for guiding retrohoming[3]. Relocating these homing structures may allow researchers to redirect DGR-mediated diversification to a heterologous position, and has been previously demonstrated by shifting retrohoming from the Mtd variable region to a kanamycin resistance reporter gene on either the replicating phage or the prophage in the bacterial chromosome[3]. This discovery opens the possibility of directing the *Bordetella* phage DGR to affect diversity generation on a protein target of choice. Additionally, Overstreet and colleagues have demonstrated that DGR also tolerates heterologous peptide insertions up to 19 bp at four distinct *mtd* positions, dramatically increasing the theoretical diversity of the Mtd library[4].

Structural characteristics of protein-protein interfaces such as residue packing, hydrophobicity, contact size, or shape complementarity have so far been largely unsuccessful at predicting binding affinity[5–7]. Thus, directed evolution continues to be an important tool in the protein engineering space and further studies to explore amenable protein scaffolds will be vital. Further downstream the protein engineering process, considerations into protein solubility and yield are central to the commercial viability of protein-based therapeutics and diagnostics. Unfortunately, protein ligands that demonstrate high-affinity binding are frequently missed or even knowingly ignored due to solubility and yield issues[8,9]. Small sequence changes or even single-site mutations often drive results in misfolded product in an inclusion body. Thus, continued advancements on both these components of the protein engineering pipeline are crucial to lowering costs to enable novel biological applications.

REFERENCES

1. Alayyoubi, M., Guo, H., Dey, S., Golnazarian, T. & Garrett, A. Structure of the essential diversity-generating retroelement protein bAvd and its functionally important interaction with reverse transcriptase. *Structure.* **21,** 266–276 (2014).

2. Jaeger, J. & Belfort, M. Hand-holding for retrohoming in a molecular diversity dance. *Structure.* **21,** 195–196 (2013).

3. Guo, H. *et al.* Target site recognition by a diversity-generating retroelement. *PLoS Genet.* **7,** e1002414 (2011).

4. Overstreet, C. M. *et al.* Self-made phage libraries with heterologous inserts in the Mtd of Bordetella bronchiseptica. *Protein Eng. Des. Sel.* **25,** 145–51 (2012).

5. Norel, R., Lin, S. L., Wolfson, H. J. & Nussinov, R. Shape complementarity at protein-protein interfaces. *Biopolymers.* **34,** 933–940 (1994).

6. Lo Conte, L., Chothia, C. & Janin, J. The atomic structure of protein-protein recognition sites. *J. Mol. Biol.* **285,** 2177–2198 (1999).

7. Gilbreth, R. N. & Koide, S. Structural Insights for Engineering Binding Proteins Based on Non-Antibody Scaffolds. *Curr. Opin. Struct. Biol.* **22,** 413–420 (2013).

8. Tsumoto, K., Ejima, D., Kumagai, I. & Arakawa, T. Practical considerations in refolding proteins from inclusion bodies. *Protein Expr. Purif.* **28,** 1–8 (2003).

9. Vallejo, L. F. & Rinas, U. Strategies for the recovery of active proteins through refolding of bacterial inclusion body proteins. *Microb. Cell Fact.* **3,** 11 (2004).