Fibronectin Domain Engineering

by

Benjamin Joseph Hackel

B.S., Chemical Engineering
University of Wisconsin-Madison, 2003

Submitted to the Department of Chemical Engineering
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY IN CHEMICAL ENGINEERING
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2009

© Massachusetts Institute of Technology 2009
All Rights Reserved

ARCHIVES
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
JUN 0 2 2010
LIBRARIES

Signature of Author: ______________________
Department of Chemical Engineering
August 13, 2009

Certified by: ______________________
K. Dane Wittrup
Professor of Chemical Engineering
Thesis Supervisor

Accepted by: ______________________
William M. Deen
Professor of Chemical Engineering
Chairman, Committee for Graduate Students

Fibronectin Domain Engineering

by

Benjamin Joseph Hackel

Submitted to the Department of Chemical Engineering on August 13, 2009 in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Chemical Engineering

ABSTRACT

Molecular recognition reagents are a critical component of targeted therapeutics, *in vivo* and *in vitro* diagnostics, and biotechnology applications such as purification, detection, and crystallization. Antibodies have served as the gold standard binding molecule because of their high affinity and specificity and, historically, because of their ability to be generated by immunization. However, antibodies suffer from several shortcomings that hinder their production and reduce their efficacy in a breadth of applications. The tenth type III domain of human fibronectin provides a small, stable, single-domain, cysteine-free protein scaffold upon which molecular recognition capability can be engineered.

In the current work, we provide substantial improvements in each phase of protein engineering through directed evolution and develop a complete platform for engineering high affinity binders based on the fibronectin domain. Synthetic combinatorial library design is substantially enhanced through extension of diversity to include three peptide loops with inclusion of loop length diversity. The efficiency of sequence space search is improved by library focusing with tailored diversity for structural bias and binding capacity. Evolution of lead clones was substantially improved through development of recursive dual mutagenesis in which each fibronectin gene is subtly mutated or the binding loops are aggressively mutated and shuffled. This engineering platform enables robust generation of high affinity binders to a multitude of targets. Moreover, the development of this technology is directly applicable to other protein engineering campaigns and advances the scientific understanding of molecular recognition.

Binders were engineered to tumor targets carcinoembryonic antigen, CD276, and epidermal growth factor receptor as well as biotechnology targets human serum albumin and goat, mouse, and rabbit immunoglobulin G. Binders have demonstrated utility in affinity purification, laboratory detection, and cellular labeling and delivery. Of particular interest, a panel of domains was engineered that bind multiple epitopes of epidermal growth factor receptor. Select non-competitive heterobivalent combinations of binders effectively downregulate receptor in a non-agonistic manner in multiple cell types. These agents inhibit proliferation and migration and provide a novel potential cancer therapy.

Thesis Supervisor: K. Dane Wittrup, C.P. Dubbs Professor of Chemical Engineering and Biological Engineering

ACKNOWLEDGEMENTS

This thesis is the result of the collective effort, insight, and support of a multitude of people, and for that I am deeply grateful.

My advisor, Dane Wittrup, has provided a superb blend of freedom and guidance throughout this project. His ability to maintain the 50,000-foot view but still appreciate the most minute, yet critical, detail of any experiment is invaluable. He also deserves praise for assembling a constructive lab community. Dasa Lipovsek initiated fibronectin domain work in the Wittrup lab and provided guidance and experimental instruction. Andy Rakestraw helped me to appreciate the impactful element of any experiment or project. Pankaj Karande, Shaun Lippow, Stefan Zajic, and Greg Thurber imparted their wisdom in helpful discussions. Shanshan Howland reminded me to progress one experiment at a time. Ginger Chao taught me about EGFR and provided experimental instruction. Steve Sazinsky offered experimental tips and conversations full of inspiration, insight, and serenity. Margie Ackerman was a source of many useful scientific and professional discussions. Annie Gai patiently answered my numerous questions. Mike Schmidt, Kelly Davis, David Liu, Chris Pirie, and John Rhoden supplied useful reagents, protocols, and discussions. Jamie Spangler took part in many useful EGFR discussions. Jordi Mata-Fink provided support, friendship, and a steady supply of witticisms.

My thesis committee, Arup Chakraborty and Bob Sauer, provided useful scientific and professional advice. Undergraduate researchers Atul Kapila and Selasie Goka provided advances in protein production techniques while Minah Shahbaz and Danielle Wang furthered protein engineering. Dasa Lipovsek provided the initial single- and double-loop fibronectin libraries. Steve Sazinsky contributed biotinylated EGFR ectodomain. Ginger Chao supplied the EGFR epitope-mapping library. Doug Lauffenburger (MIT) provided the autocrine signaling cell line. Jeffrey Ravetch (Rockefeller University) provided Fc receptors.

I am also grateful for the support of friends and family. My parents taught me the importance of hard work and my brothers, Mike and Scott, unwittingly provided a constant stream of goals for which to aim.

To my son Jordan, a constant source of smiles as I concluded this thesis. To my daughter Adyson, whose hugs and love were not swayed by the success, or lack thereof, of each day's experiments. To my wife, Mary, whose unbelievable support and understanding has strengthened me throughout. Your encouragement and sacrifice deserve far more than words can offer. Thank you.

TABLE OF CONTENTS

1. Background 7

Protein Scaffolds 7
Protein Engineering 11
Epidermal Growth Factor Receptor 14
Thesis Overview 18
References 19

2. Evolution through Recursive Mutagenesis and Loop Shuffling . . . 22

Introduction 22
Results 25
Discussion 40
Materials and Methods 44
References 52

3. Constrained Protein Diversity with a Tyrosine/Serine Code and DE Loop Bias 55

Introduction 55
Results 56
Discussion 65
Materials and Methods 67
References 72

4. Stability and Complementarity Bias Improve the Protein Functionality Landscape 74

Introduction 74
Results 75
Discussion 89
Materials and Methods 90
References 95

5. Epidermal Growth Factor Receptor Downregulation with Bivalent Fibronectin Constructs 97

Introduction 97
Results 98
Discussion 117
Materials and Methods 120
References 128

6. Concluding Remarks 130

Appendix A. Error-prone PCR Modeling 134
Appendix B. Engineering CEA Binders for Tumor Targeting 146
Appendix C. Engineering CD276 Binders for Targeting Tumor Vasculature . 153
Appendix D. DNA Sequences 158
Appendix E. Protocols 172
Curriculum Vitae 183

LIST OF ABBREVIATIONS

CDR	complementarity-determining region
CEA	carcinoembryonic antigen
CHO	Chinese hamster ovary
DMEM	Dulbecco's modified Eagle medium
EGF	epidermal growth factor
EGFR	epidermal growth factor receptor
ERK	extracellular signal-regulated kinase
FACS	fluorescence-activated cell sorting
FBS	fetal bovine serum
FcγR	Fcγ receptor
Fn3	tenth type III domain of human fibronectin
HEK	human embryonic kidney
HSA	human serum albumin
IgG	immunoglobulin G
K_d	equilibrium dissociation constant
mIgG	mouse immunoglobulin G
PBSA	phosphate-buffered saline with bovine serum albumin
T_m	midpoint of thermal denaturation
VEGF-R2	vascular endothelial growth factor receptor 2

1. BACKGROUND

Protein Scaffolds

Molecules capable of specific, high affinity binding to a molecular target are vital to targeted therapeutics, *in vivo* and *in vitro* diagnostics, and biotechnology applications such as purification, detection, and crystallization. Antibodies have been the binding reagent of choice because of their binding capacity and ability to be generated naturally by immunization. However, antibodies suffer from several shortcomings that reduce their efficacy in a breadth of applications and hinder their production.

Full immunoglobulins are large (~150 kDa), glycosylated molecules comprising multiple domains connected by hydrophobic interactions and disulfide bonds. The reduced diffusivity resulting from its large size impedes *in vivo* delivery, such as solid tumor uptake.[1] The large size and neonatal Fc receptor interaction slow clearance, which is beneficial for sustained therapeutic use, but detrimental for *in vivo* imaging. Size also limits the effectiveness of antibodies for Förster resonance energy transfer by separating the binding site from the fluorophore. Glycosylation, disulfide bonds, and multi-domain structure complicate production, which is generally conducted in expensive mammalian cell culture. Structural complexity hinders protein fusion such as for bispecific formats or display technologies for protein engineering. Moreover, the presence and importance of disulfide bonds generally precludes use in the reducing intracellular environment. Multiple cysteines also complicate thiol-based conjugation, which would otherwise allow simple inclusion of reporter moieties for diagnostics, immobilization for biotechnology applications, or attachment of therapeutic payloads such as radionuclides, toxins, or proteases. Antibodies also suffer from moderate instability. Resultant degradation and aggregation of *in vivo* diagnostics and therapeutics reduces potency and can elicit an immune response. Antibody shelf-life is suboptimal, reducing efficacy and elevating costs. Instability also limits the robustness of antibodies in biotechnology applications such as the stringent washing steps of purification and detection.

The opportunity for improvement over antibodies has not evaded the scientific community as a multitude of alternative scaffolds have been investigated. Molecular recognition scaffolds are conserved structural frameworks tolerant of targeted amino acid variability enabling diverse binding function, and thus include antibodies and their subdomains. However, alternative scaffolds can be smaller, single domains, more stable, cysteine-free, more easily produced, simpler to engineer, or tailored for a specific application. Numerous scaffolds have been investigated;[2; 3] a few of the more successful and well-characterized are highlighted here (Figure 1.1).

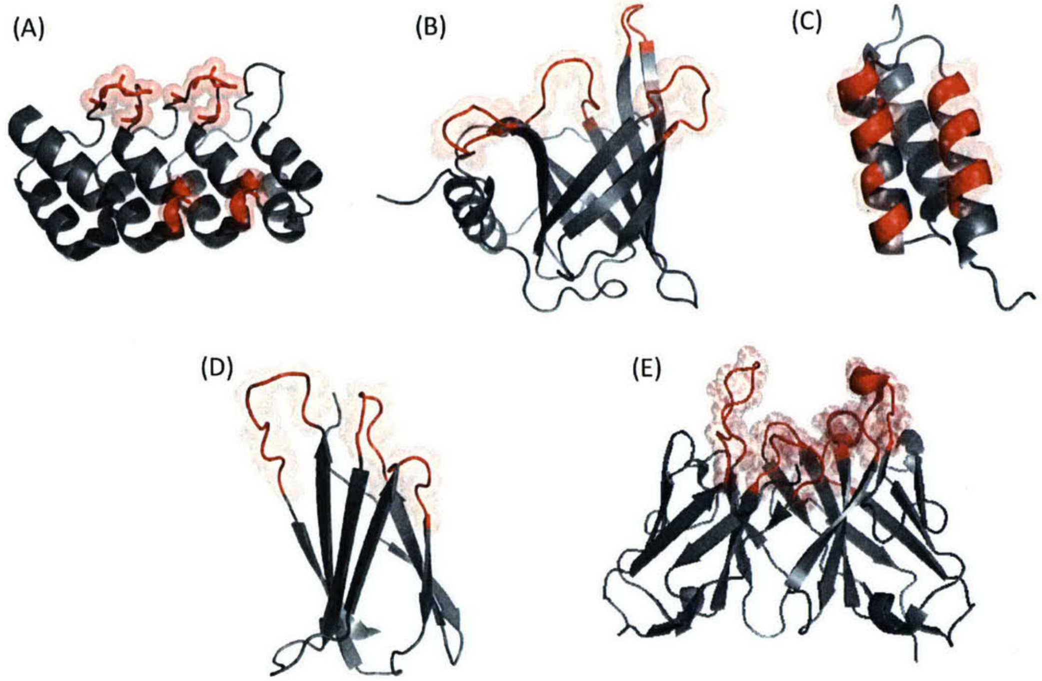


Figure 1.1 *Protein scaffolds for molecular recognition.* (A) Designed ankyrin repeat protein (PDB ID: 2JAB). (B) Anticalin (PDB ID: 1RBP). (C) Affibody (PDB ID: 2B88). (D) Fibronectin domain (PDB ID: 1TTG) (E) Single-chain antibody fragment (PDB ID: 1X9Q). Residues that are generally diversified to generate novel binding function are depicted in red with semi-transparent spheres. Images were created in MacPymol.

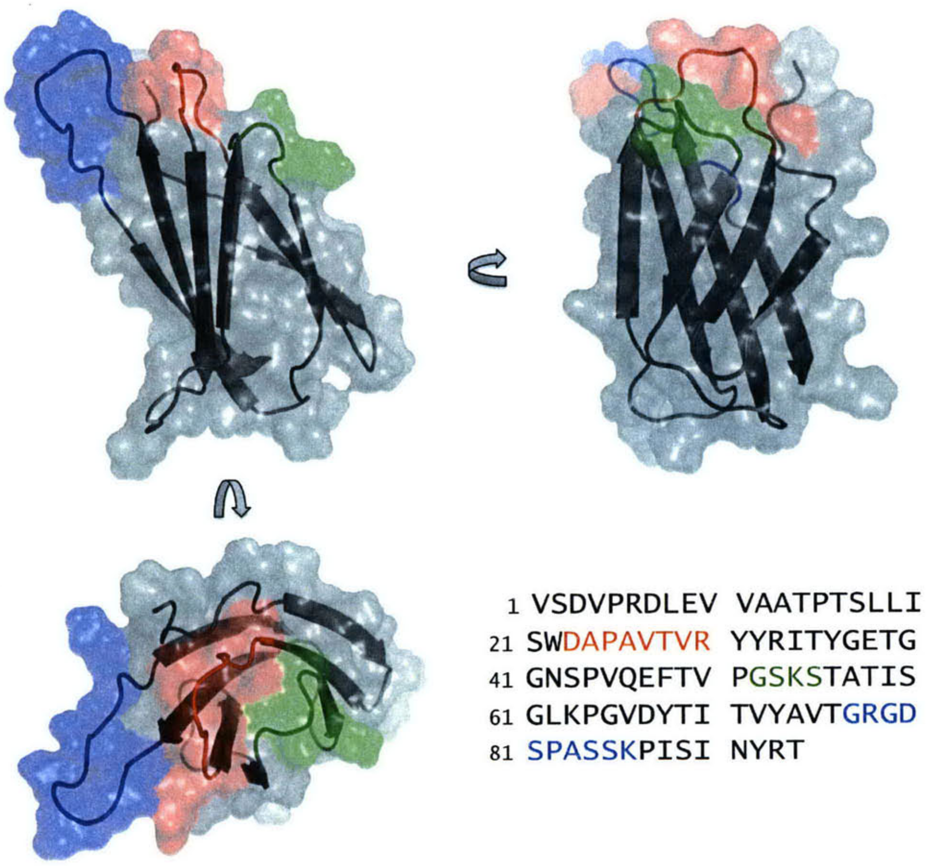
Designed ankyrin repeat proteins are composed of several 33 amino acid modules that each forms a β-turn, antiparallel α-helices, and a loop (Figure 1.1A).[4] The cysteine-free proteins have high thermodynamic stability (~10 kcal/mol), are produced at 200 mg/L in *E. coli*, and are readily crystallized. Targeted diversification of surface residues along with mutation during ribosome display screening has yielded nanomolar to picomolar binders to a variety of protein targets. Although their fully synthetic origins raise concerns about immunogenicity for *in vivo* use, early experimental evidence suggests that this concern may not be substantiated.

Anticalins, based on natural ligand-binding lipocalins, are 160-180 amino acid β-barrels with a pocket formed by four loops (Figure 1.1B).[5] The concave binding pocket is best served for hapten binding yet variation in the four loops has yielded nanomolar binders to multiple targets including digitoxigenin and cytotoxic T lymphocyte antigen 4. On the down side, the scaffold contains several disulfide bonds, is only moderately stable (~6 kcal/mol), and has low bacterial expression (0.5 mg/L).

Another scaffold is the affibody, a 58 amino acid, three-helix bundle derived from the Z domain of staphylococcal protein A (Figure 1.1C).[6] Diversification of thirteen residues on the surface of two antiparallel helices has yielded binders to multiple targets; however, binders are characterized by high association and dissociation rates generally resulting in micromolar affinity, though mid-nanomolar binders have been identified. The cysteine-free domains are produced in *E. coli* at up to 200 mg/L, but instability has been problematic including the presence of molten globules with a 30% reduction in helicity and 37° thermal destabilization relative to wild-type.

An additional scaffold under study is based on the tenth type III domain of human fibronectin (Fn3). The scaffold is small (94 amino acids, 10 kDa), stable (7.5-9.4 kcal/mol, T_m = 90°C),[7; 8] soluble to 15 mg/mL, free of cysteines and expressed at ~50 mg/L in *E. coli*.[9] Depending on the degree of modification, it is reasonable to expect low immunogenicity *in vivo* due to stability and natural abundance, as the Fn3 domain occurs in ~2% of animal proteins.[10] In addition, both solution[11] and crystal[12] structures of Fn3 have been determined enabling rational elements of design. The scaffold contains three solvent-exposed loops on either side of parallel β -sheets, somewhat akin to the immunoglobulin fold (Figure 1.2). Significant evidence exists that Fn3 loops can tolerate diversity to potentially function in analogous fashion to complementarity-determining regions of antibodies. Sequence analyses reveal large variations in the BC and FG loops (Fn3 loops can be referenced by the two peripheral β-strands) with moderate variation in

DE loop sequences. NMR spectroscopy indicates significant flexibility of the FG loop as well as moderate flexibility of BC.[13] Moreover, elongation by insertion of four glycine residues is moderately well tolerated (1.2, 2.3, and 0.4 kcal/mol destabilization of BC, DE, and FG).[14] The opposing loops, AB, CD, and EF, offer potential for a bispecific scaffold but are neither as well arranged nor as tolerable of insertion as the other loops. In short, Fn3 provides numerous biophysical advantages over antibodies and is an attractive scaffold for engineering binding proteins.



1 VSDVPRDLEV VAATPTSLLI
21 SWDAPAVTVR YYRITYGETG
41 GNSPVQEFTV PGSKSTATIS
61 GLKPGVDYTI TVYAVTGRGD
81 SPASSKPISI NYRT

Figure 1.2 *Fibronectin domain.* The solution structure (PDB ID: 1TTG) of Fn3 is presented with 90° rotations. The wild-type sequence is indicated. The BC (red), DE (green), and FG (blue) loops are highlighted.

Use of this scaffold as a binding protein has been proven in several applications. The native scaffold is a natural binder as the FG loop contains the Arg-Gly-Asp tripeptide critical in integrin binding.[15] In the initial use of the domain as a scaffold for molecular recognition, randomization of the BC loop and a shortened FG loop yielded micromolar binders to ubiquitin.[16] This study demonstrated the ability of Fn3 to accommodate mutations in loop residues without notable structural change and the ability to introduce novel binding function. However, reduced stability and solubility and non-specific, low affinity binding characterized the Fn3 variant. Screening of a library with more extensive randomization of the BC, DE, and FG loops yielded binders to tumor necrosis factor α and vascular endothelial growth factor receptor 2 (VEGF-R2) of nanomolar affinity.[8; 9] Further maturation resulted in binders of sub-nanomolar affinity demonstrating the potential for high affinity binding with Fn3. Engineered Fn3 variants have been used intracellularly,[17] as inhibitors in cell culture,[18] in protein arrays,[9] and as labeling reagents in flow cytometry[18] and Western blots.[19] An anti-VEGF-R2 Fn3 is progressing through clinical trials.

Fn3 shows great promise as a scaffold for molecular recognition and has been substantiated in a variety of cases. However, reported binding variants are generally characterized by substandard biophysical parameters. More thorough understanding of the protein, both in terms of inherent biophysics and mode of molecular recognition, as well as improved search of scaffold sequence space should advance Fn3 from isolated successes to a robust scaffold for molecular recognition therapeutics and reagents. Scaffold development should provide broad impact into protein engineering technology and further the understanding of molecular recognition and protein sequence/function relationships.

Protein Engineering

Engineering novel binding proteins requires identification of a functional primary protein sequence from the enormous number of possible sequences (*e.g.*, for a protein the size of Fn3, there are $20^{94} = 10^{122}$ possible proteins). Neither sequence/structure nor structure/function relationships are understood well enough to enable robust *de novo* theoretical design; thus, screening and evolution of combinatorial libraries provides the most effective route to protein engineering of novel binders. This process involves three key elements: *naïve* library design, selection of functional clones, and sequence diversification of lead clones (Figure 1.3).

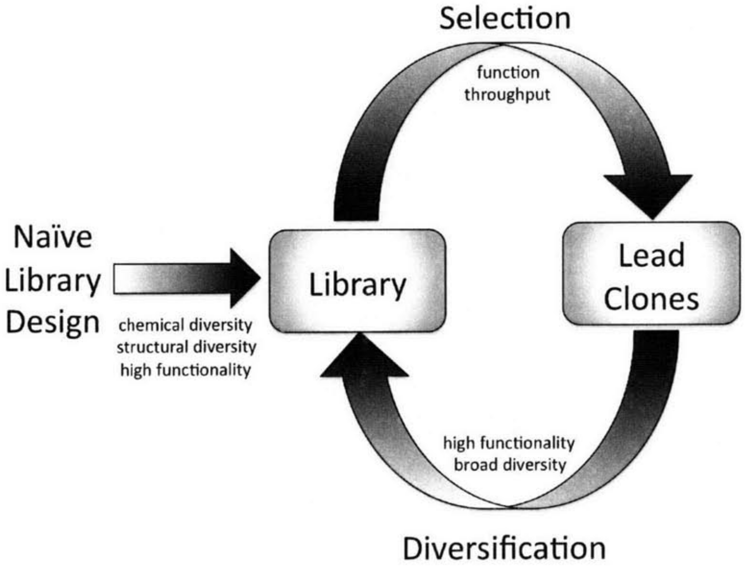

Selection
function
throughput
Naïve
Library
Design
chemical diversity
structural diversity
high functionality
Library
Lead
Clones
high functionality
broad diversity
Diversification

Figure 1.3 *Protein engineering by directed evolution.* A naïve combinatorial library of protein clones undergoes selection or screening to identify the lead clones. These clones are diversified via mutagenesis or informed library synthesis to yield a next generation library. This cycle continues until the desired functionality is achieved.

Naïve library design requires the balance of creating sufficient conformational and chemical diversity to yield high affinity binding to myriad epitopes while providing a high frequency of functional clones such that the limited searchable sequence space contains clones with the appropriate phenotype. The essence of library design is summarized in two questions: In which amino acid positions should diversity be included? Which amino acids should be included at those positions?

Clones with the desired functionality can be identified from the library of protein variants with high throughput selection via linkage of genotype and phenotype. Though this linkage can be achieved through a multitude of display formats[20] such as phage display and mRNA display, yeast surface display is the superior method.[21] *In vitro* technologies tout high theoretical library sizes because of the absence of cellular transformation, which can limit library size. Yet in a recent comparison of yeast surface display and phage display using the same antibody DNA library and target antigen, yeast surface display identified three times more clones than did phage display and did not miss a single phage clone revealing that constructed size and functional size can differ substantially.[22] Yeast surface display may also enable selection of stable clones because of the quality control apparatus of the eukaryotic secretory system.[23] Fluorescence-activated cell sorting of yeast allows quantitative discrimination of clone functionality.[24]

In yeast surface display, tens of thousands of copies of Fn3 are tethered to the exterior of an individual *Saccharomyces cerevisiae* yeast cell while the genetic information for the Fn3 clone is maintained in the cell interior. The cell-protein linkage begins with the Aga1p subunit of a-agglutinin, which anchors in the cell wall periphery via β-glucan covalent linkage.[25] The Aga2p subunit, secreted from the yeast cell as a fusion to Fn3, attaches to Aga1p via two disulfide bonds. The peptide bond in the fusion protein thus completes the linkage resulting in 'display' of Fn3 on the yeast cell (Figure 1.4A). Aga2p and Fn3, linked by a $(G_4S)_3$ peptide, are followed by HA and c-myc epitopes, respectively, to enable analysis of the display of Aga2p and the full-length protein fusion. Display is achieved through transformation of DNA encoding for the Aga2p-Fn3 fusion (Figure 1.4B) followed by cell growth and induction of both Aga1p and Aga2p-Fn3 protein expression using a galactose-inducible GAL promoter. The displayed clones can be screened for their ability to bind to a target of interest using flow cytometry or captured by immobilized antigen.

Selected clones can be evolved through partial diversification of their sequence followed by selection for mutants that exhibit improved functionality. Error-prone PCR to introduce random mutations throughout the gene is the most common method of diversification. Yeast surface display also enables gene shuffling via homologous recombination.[26]

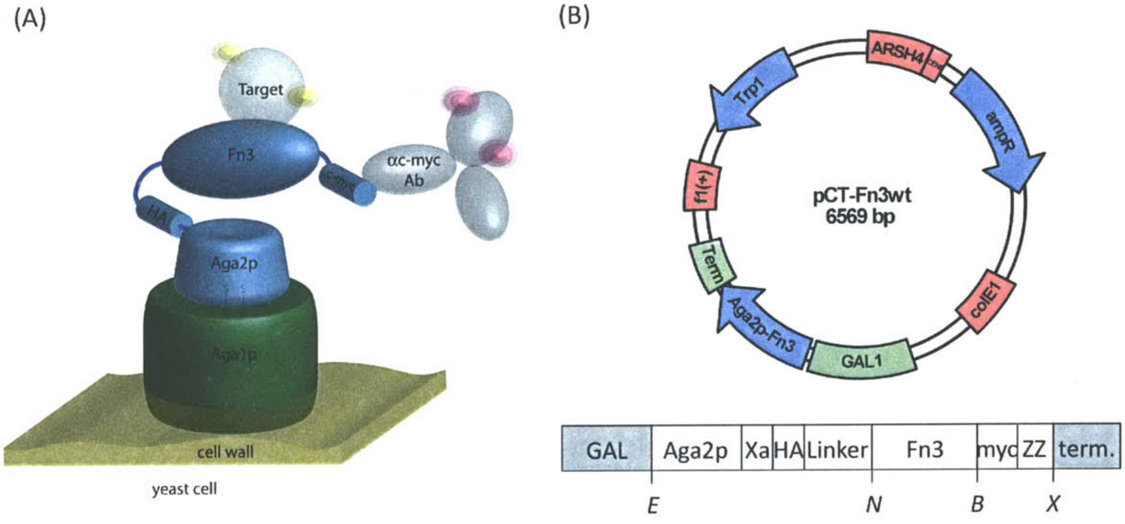
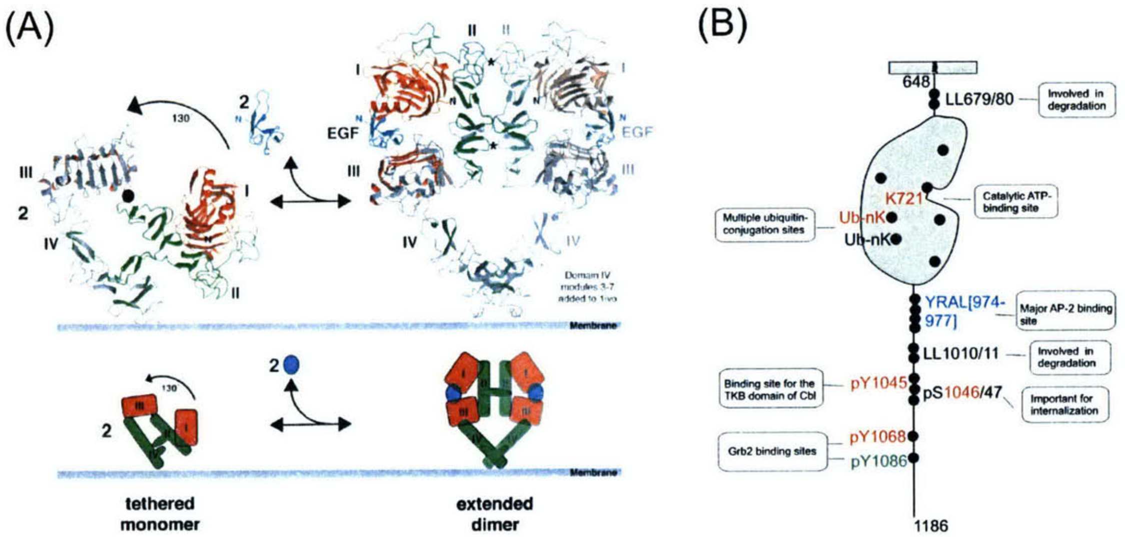


Figure 1.4. *Yeast surface display.* (A) Schematic of yeast surface display. The Aga2p subunit is tethered to the yeast cell via two disulfide bonds to Aga1p, which is anchored in the yeast cell wall. The protein of interest is connected to the Aga2p subunit through a flexible peptide linker as a result of genetic fusion. HA and c-myc epitope tags flank the protein of interest on the N- and- C-termini enabling detection of each displayed molecule using fluorophore-conjugated antibodies. Depending on the protein of interest and induction conditions, tens of thousands of these fusions are displayed per cell. (B) pCT-Fn3wt Vector. *ARSH4*: autonomously replicating sequence H4; *CEN6*: centromeric sequence 6; *ampR*: ampicillin resistance gene; *colE1*: colicin E1 origin of replication; *GAL1*: GAL1 promoter; *E*: EcoRI site; *Aga2p*: agglutinin 2p protein subunit; *Xa*: factor Xa cleavage site; *HA*: hemagluttinin epitope; *linker*: $(Gly_4Ser)_3$ linker; *N*: NheI site; *Fn3*: fibronectin tenth type III gene; *B*: BamHI site; *myc*: c-myc epitope; *ZZ*: TAATAG stop codons; *X*: XhoI site; *term*: alpha mating factor terminator; *f1(+)*: f1 origin of replication; *Trp1*: Trp1 gene. The lower representation is not to scale.

Epidermal Growth Factor Receptor

The epidermal growth factor receptor (EGFR) is a receptor tyrosine kinase in the ErbB family. The receptor comprises three regions: extracellular, transmembrane, and intracellular, which consists of a juxtamembrane domain, kinase domain, and a C-terminal tail containing phosphorylation sites (Figure 1.5A,B). The extracellular region consists of four domains of which domains I and III are leucine rich repeat folds and domains II and IV are cysteine-rich domains. The receptor is predominantly present in a tethered conformation on the cell surface. Binding of ligand, including epidermal growth factor, transforming growth factor α, epiregulin, amphiregulin, β-cellulin, and heparin-binding epidermal growth factor, stabilizes an open conformation of the receptor. Resultant dimerization enables kinase activation and phosphorylation of the intracellular domain. Phosphorylation sites enable docking of adaptor proteins that initiate signaling cascades such as the mitogen-activated protein kinase pathway activated by Ras and Shc, the Akt pathway activated by phosphatidylinositol-3-OH kinase, and the protein kinase C pathway activated by phospholipase Cγ. These pathways form a complex signaling network that impacts multiple cellular processes including differentiation, migration, and growth[27] (Figure 1.5C). Activated EGFR is endocytosed within several minutes and a fraction undergoes fast recycling from the early endosome. The alternate fraction persists to the late endosome resulting in slower recycling or degradation (Figure 1.5D).[28]


(A)
130
2
III
2
IV
I
II
EGF
II
II
I
I
EGF
III
III
IV
IV
Domain IV modules 3-7 added to 1ivo
Membrane
2
130
2
III
I
Membrane
tethered monomer
extended dimer
(B)
648
LL679/80
Involved in degradation
K721
Catalytic ATP-binding site
Multiple ubiquitin-conjugation sites
Ub-nK
Ub-nK
YRAL[974-977]
Major AP-2 binding site
LL1010/11
Involved in degradation
Binding site for the TKB domain of Cbl
pY1045
pS1046/47
Important for internalization
Grb2 binding sites
pY1068
pY1086
1186

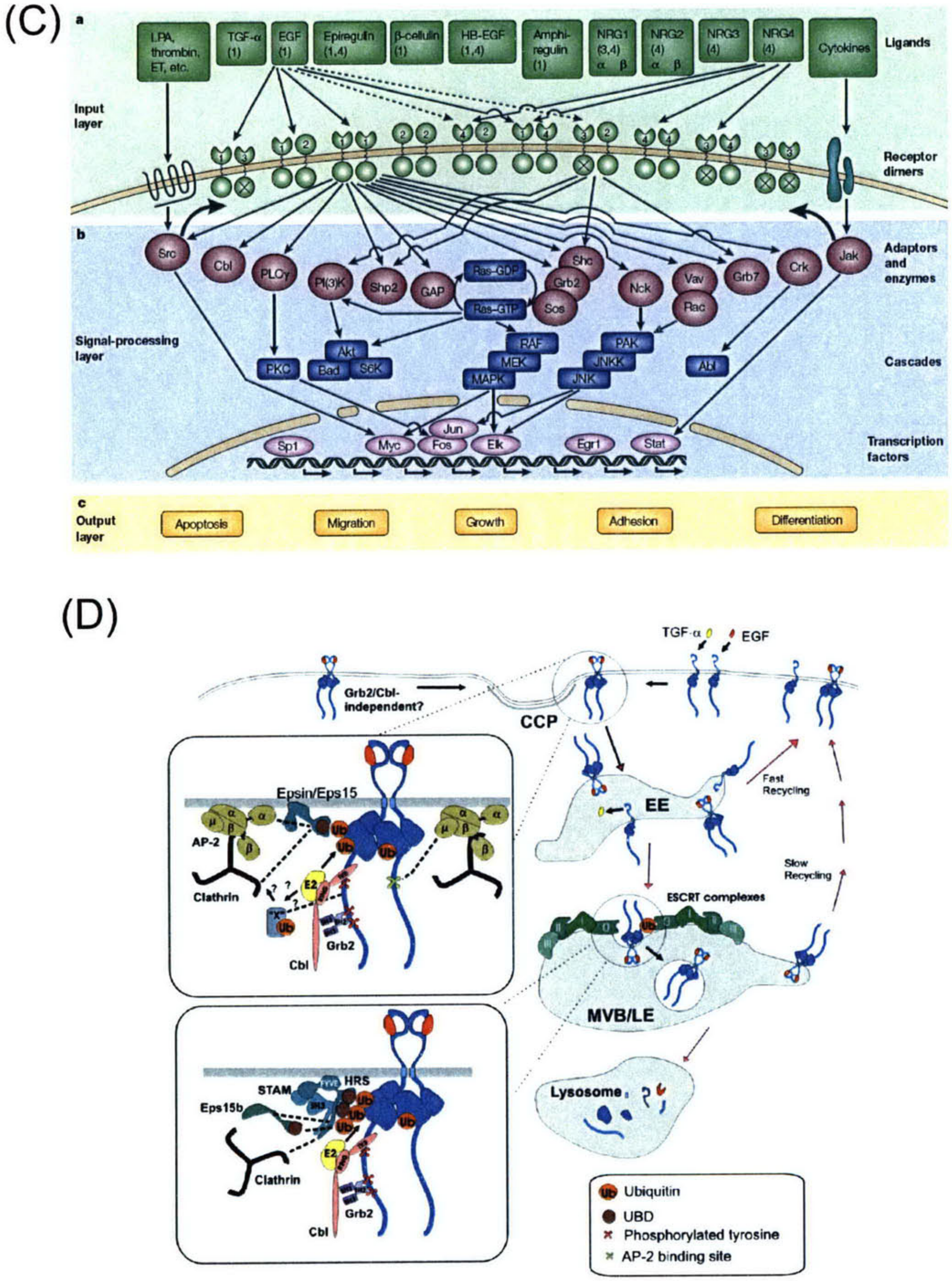

(C)
a
LPA, thrombin, ET, etc.
TGF-α (1)
EGF (1)
Epiregulin (1,4)
β-cellulin (1)
HB-EGF (1,4)
Amphi-regulin (1)
NRG1 (3,4) α β
NRG2 (4) α β
NRG3 (4)
NRG4 (4)
Cytokines
Ligands
Input layer
Receptor dimers
b
Src
Cbl
PLCγ
PI(3)K
Shp2
GAP
Ras–GDP
Ras–GTP
Shc
Grb2
Sos
Nck
Vav
Rac
Grb7
Crk
Jak
Adaptors and enzymes
Signal-processing layer
Akt
Bad
S6K
PKC
RAF
MEK
MAPK
PAK
JNKK
JNK
Abl
Cascades
Jun
Sp1
Myc
Fos
Elk
Egr1
Stat
Transcription factors
c
Output layer
Apoptosis
Migration
Growth
Adhesion
Differentiation
(D)
TGF-α
EGF
Grb2/Cbl-independent?
CCP
Epsin/Eps15
AP-2
Clathrin
E2
Ub
Grb2
Cbl
EE
Fast Recycling
Slow Recycling
ESCRT complexes
MVB/LE
Lysosome
STAM
HRS
FYVE
Eps15b
Clathrin
E2
Grb2
Cbl
Ubiquitin
UBD
Phosphorylated tyrosine
AP-2 binding site

Figure 1.5. *EGFR structure and function.* (A) The extracellular region of EGFR, consisting of four domains, switches from a closed to open conformations upon EGF binding. Figure reproduced.[29] (B) The juxtamembrane, kinase, and phosphorylation tail domains of the EGFR intracellular region are shown. Key phosphorylation sites are highlighted. Figure reproduced.[28] (C) A systems-level schematic of ErbB signaling is depicted. Figure reproduced.[27] (D) EGFR trafficking is detailed. Figure reproduced.[28]

Given the impact of EGFR on a variety of cellular processes, it is perhaps predictable that aberrant signaling from the receptor is implicated in multiple cancers. Dysregulation is observed in breast, bladder, head and neck, and non-small cell lung cancers.[27] An analysis of 15 years of published literature on EGFR expression and cancer prognosis revealed that receptor overexpression is associated with reduced survival in 70% of head and neck, ovarian, cervical, bladder, and esophageal cancers.[30] Autocrine production of transforming growth factor α and epidermal growth factor (EGF) correlate with reduced survival in lung cancer.[31] Receptor mutation is also implicated in cancer. EGFRvIII, which lacks amino acids 6-273, is observed in glioblastoma, non-small cell lung cancer, and cancers of the breast and ovary.[32] This mutant is unable to bind ligand yet is constitutively active, posing a unique therapeutic challenge, particularly for ligand blocking agents. Ectodomain point mutants in glioblastoma yield tumorigenicity.[33] Kinase domain mutations observed in non-small cell lung cancer hyperactivate kinase.[34]

As a result of the involvement of EGFR in cancer, there has been substantial effort spent developing receptor inhibitors as therapeutics. The U.S. Food and Drug Administration has approved two monoclonal antibodies and two tyrosine kinase inhibitors targeting EGFR. Cetuximab (Erbitux, Bristol-Myers Squibb), approved for colorectal and head and neck cancer, and panitumumab (Vectibix, Amgen), approved for colorectal cancer, are antibodies that compete with EGF for receptor binding. However, the relative impact of ligand competition, receptor downregulation, and antibody-dependent cellular cytotoxicity is unknown (note that panitumumab is an immunoglobulin G (IgG) 2a molecule and thus incapable of triggering cellular cytotoxicity). Both antibodies exhibit modest efficacy. In treatment of metastatic colorectal cancer refractory to irinotecan tyrosine kinase inhibitor, only 11% of patients respond to cetuximab alone and only 23% respond to cetuximab and irinotecan in combination.[35] In the treatment of head and neck cancer, the addition of cetuximab to radiation extends median survival from 29 to 49 months yet only increases responsiveness from 45% to 55% and improvement is only evident for oropharyngeal cancer but not hypopharyngeal or laryngeal cancers. Moreover, metastases were present at comparable amounts with and without antibody.[36] In metastatic colorectal cancer, panitumumab extends progression-free survival from 64 days to 90 days; yet the overall response rate was only 8% and there was no improvement in overall survival.[37]

While this efficacy validates EGFR as a useful therapeutic target, it begs the search for improved understanding of receptor biology and the development of improved therapy. Potential causes of the modest efficacy include inability to effectively compete with ligand, especially in the presence of autocrine signaling; insufficient downregulation of receptor; lack of inhibition of constitutively active EGFRvIII; and mutational escape. Thus, novel binders capable of downregulation and/or inhibition via different modes of action would be beneficial. Small, monovalent binders would enable improved biophysical studies via specific inhibition or Forster resonance energy transfer. Such small binders could also be useful for *in vivo* imaging to study receptor localization and trafficking.

Thesis Overview

In this work, a platform for engineering Fn3 domains as binding reagents was developed. The efficacy and efficiency of the maturation of lead clones was vastly improved (Chapter 2). Parallel weak mutagenesis of the entire gene and strong mutagenesis of the loops with forced shuffling enhanced both the breadth and quality of sequence diversification. The quality was further enriched through focusing of error-prone PCR on ideal conditions identified by mathematical modeling. Naïve library design progressed through improvement of multiple elements: inclusion of loop length diversity (introduced in Chapter 2 and substantiated in Chapters 3 and 4), investigation of constrained amino acid diversity (Chapter 3), design and validation of tailored amino acid diversity for complementarity and stability (Chapter 4). In addition, though absent from this thesis since the project was led by colleague Margaret Ackerman, a magnetic bead selection method was developed to improve isolation of low affinity binders as well as selection throughput and removal of reagent binders.[38] In total all key elements of protein engineering by directed evolution were improved to produce a platform enabling the rapid and robust isolation of high affinity binders from the Fn3 scaffold. Importantly, these technological developments are broadly applicable to the field of protein engineering.

The final platform has enabled isolation of binders to eight targets: cancer targets human EGFR, human A33, mouse A33, and mouse CD276; immunological targets Fcγ receptors IIa and IIIa; and biotechnological targets mouse IgG and human serum albumin (HSA). In addition, binders to lysozyme, carcinoembryonic antigen, goat IgG, and rabbit IgG were engineered during platform development. Collectively, these binders represent useful reagents for biotechnology, scientific, and clinical applications.

Most notably, EGFR binders were incorporated into a novel bispecific format. Selective non-competitive heterobivalent constructs are capable of receptor downregulation. Select constructs inhibit cell proliferation and migration, particularly in combination with a ligand-competitive antibody, and therefore have strong therapeutic potential (Chapter 5).

REFERENCES

1. Thurber, G. & Wittrup, K. D. (2008). Quantitative spatiotemporal analysis of antibody fragment diffusion and endocytic consumption in tumor spheroids. *Cancer Research* **68**, 3334-41.
2. Binz, H., Amstutz, P. & Plückthun, A. (2005). Engineering novel binding proteins from nonimmunoglobulin domains. *Nat Biotechnol* **23**, 1257-1268.
3. Skerra, A. (2007). Alternative non-antibody scaffolds for molecular recognition. *Current Opinion in Biotechnology* **18**, 295-304.
4. Binz, H., Amstutz, P., Kohl, A., Stumpp, M., Briand, C., Forrer, P., Grütter, M. & Plückthun, A. (2004). High-affinity binders selected from designed ankyrin repeat protein libraries. *Nat Biotechnol* **22**, 575-582.
5. Skerra, A. (2008). Alternative binding proteins: anticalins - harnessing the structural plasticity of the lipocalin ligand pocket to engineer novel binding activities. *FEBS J* **275**, 2677-83.
6. Nygren, P. A. (2008). Alternative binding proteins: affibody binding proteins developed from a small three-helix bundle scaffold. *FEBS J* **275**, 2668-76.
7. Cota, E. & Clarke, J. (2000). Folding of beta-sandwich proteins: three-state transition of a fibronectin type III module. *Protein Sci* **9**, 112-20.
8. Parker, M., Chen, Y., Danehy, F., Dufu, K., Ekstrom, J., Getmanova, E., Gokemeijer, J., Xu, L. & Lipovsek, D. (2005). Antibody mimics based on human fibronectin type three domain engineered for thermostability and high-affinity binding to vascular endothelial growth factor receptor two. *Protein Engineering Design and Selection* **18**, 435-444.
9. Xu, L., Aha, P., Gu, K., Kuimelis, R. G., Kurz, M., Lam, T., Lim, A. C., Liu, H., Lohse, P. A., Sun, L., Weng, S., Wagner, R. W. & Lipovsek, D. (2002). Directed evolution of high-affinity antibody mimics using mRNA display. *Chemistry & Biology* **9**, 933-42.
10. Bork, P. & Doolittle, R. F. (1992). Proposed acquisition of an animal protein domain by bacteria. *Proc Natl Acad Sci USA* **89**, 8990-4.

11. Main, A. L., Harvey, T. S., Baron, M., Boyd, J. & Campbell, I. D. (1992). The three-dimensional structure of the tenth type III module of fibronectin: an insight into RGD-mediated interactions. *Cell* **71**, 671-8.
12. Dickinson, C. D., Veerapandian, B., Dai, X. P., Hamlin, R. C., Xuong, N. H., Ruoslahti, E. & Ely, K. R. (1994). Crystal structure of the tenth type III cell adhesion module of human fibronectin. *Journal of Molecular Biology* **236**, 1079-92.
13. Carr, P. A., Erickson, H. P. & Palmer, A. G. (1997). Backbone dynamics of homologous fibronectin type III cell adhesion domains from fibronectin and tenascin. *Structure* **5**, 949-59.
14. Batori, V., Koide, A. & Koide, S. (2002). Exploring the potential of the monobody scaffold: effects of loop elongation on the stability of a fibronectin type III domain. *Protein Eng* **15**, 1015-20.
15. Pierschbacher, M. D., Hayman, E. G. & Ruoslahti, E. (1985). The cell attachment determinant in fibronectin. *J Cell Biochem* **28**, 115-26.
16. Koide, A., Bailey, C. W., Huang, X. & Koide, S. (1998). The fibronectin type III domain as a scaffold for novel binding proteins. *Journal of Molecular Biology* **284**, 1141-51.
17. Koide, A., Abbatiello, S., Rothgery, L. & Koide, S. (2002). Probing protein conformational changes in living cells by using designer binding proteins: application to the estrogen receptor. *Proc Natl Acad Sci USA* **99**, 1253-8.
18. Richards, J., Miller, M., Abend, J., Koide, A., Koide, S. & Dewhurst, S. (2003). Engineered fibronectin type III domain with a RGDWXE sequence binds with enhanced affinity and specificity to human alphavbeta3 integrin. *Journal of Molecular Biology* **326**, 1475-88.
19. Karatan, E., Merguerian, M., Han, Z., Scholle, M. D., Koide, S. & Kay, B. K. (2004). Molecular recognition properties of FN3 monobodies that bind the Src SH3 domain. *Chemistry & Biology* **11**, 835-44.
20. Hoogenboom, H. (2005). Selecting and screening recombinant antibody libraries. *Nat Biotechnol* **23**, 1105-1116.
21. Hackel, B. & Wittrup, K. (2009). Yeast Surface Display in Protein Engineering and Analysis. In *Protein Engineering Handbook* (Bronscheuer, S. L. a. U., ed.), Vol. 1. Wiley-VCH.
22. Bowley, D., Labrijn, A., Zwick, M. & Burton, D. (2007). Antigen selection from an HIV-1 immune antibody library displayed on yeast yields many novel antibodies compared to selection from the same library displayed on phage. *Protein Engineering Design and Selection* **20**, 81-90.
23. Shusta, E. V., Kieke, M. C., Parke, E., Kranz, D. M. & Wittrup, K. D. (1999). Yeast polypeptide fusion surface display levels predict thermal stability and soluble secretion efficiency. *Journal of Molecular Biology* **292**, 949-56.
24. VanAntwerp, J. J. & Wittrup, K. D. (2000). Fine affinity discrimination by yeast surface display and flow cytometry. *Biotechnol Prog* **16**, 31-7.
25. Lu, C. F., Montijn, R. C., Brown, J. L., Klis, F., Kurjan, J., Bussey, H. & Lipke, P. N. (1995). Glycosyl phosphatidylinositol-dependent cross-linking of alpha-agglutinin and beta 1,6-glucan in the Saccharomyces cerevisiae cell wall. *J Cell Biol* **128**, 333-40.
26. Swers, J. S., Kellogg, B. A. & Wittrup, K. (2004). Shuffled antibody libraries created by in vivo homologous recombination and yeast surface display. *Nucleic Acids Research* **32**, e36.
27. Yarden, Y. & Sliwkowski, M. X. (2001). Untangling the ErbB signalling network. *Nat Rev Mol Cell Biol* **2**, 127-37.
28. Sorkin, A. & Goh, L. K. (2009). Endocytosis and intracellular trafficking of ErbBs. *Experimental Cell Research* **315**, 683-696.
29. Burgess, A., Cho, H. S., Eigenbrot, C., Ferguson, K. M., Garrett, T. P., Leahy, D. J., Lemmon, M. A., Sliwkowski, M. X., Ward, C. W. & Yokoyama, S. (2003). An open-and-shut case? Recent insights into the activation of EGF/ErbB receptors. *Molecular Cell* **12**, 541-52.
30. Nicholson, R. I., Gee, J. M. & Harper, M. E. (2001). EGFR and cancer prognosis. *Eur J Cancer* **37 Suppl 4**, S9-15.
31. Tateishi, M., Ishida, T., Mitsudomi, T., Kaneko, S. & Sugimachi, K. (1990). Immunohistochemical evidence of autocrine growth factors in adenocarcinoma of the human lung. *Cancer Research* **50**, 7077-80.
32. Pedersen, M. W., Meltorn, M., Damstrup, L. & Poulsen, H. S. (2001). The type III epidermal growth factor receptor mutation. Biological significance and potential target for anti-cancer therapy. *Ann Oncol* **12**, 745-60.

33. Lee, J. C., Vivanco, I., Beroukhim, R., Huang, J. H., Feng, W. L., DeBiasi, R. M., Yoshimoto, K., King, J. C., Nghiemphu, P., Yuza, Y., Xu, Q., Greulich, H., Thomas, R. K., Paez, J. G., Peck, T. C., Linhart, D. J., Glatt, K. A., Getz, G., Onofrio, R., Ziaugra, L., Levine, R. L., Gabriel, S., Kawaguchi, T., O'Neill, K., Khan, H., Liau, L. M., Nelson, S. F., Rao, P. N., Mischel, P., Pieper, R. O., Cloughesy, T., Leahy, D. J., Sellers, W. R., Sawyers, C. L., Meyerson, M. & Mellinghoff, I. K. (2006). Epidermal growth factor receptor activation in glioblastoma through novel missense mutations in the extracellular domain. *PLoS Med* **3**, e485.
34. Sharma, S. V., Bell, D. W., Settleman, J. & Haber, D. A. (2007). Epidermal growth factor receptor mutations in lung cancer. *Nat Rev Cancer* **7**, 169-81.
35. Cunningham, D., Humblet, Y., Siena, S., Khayat, D., Bleiberg, H., Santoro, A., Bets, D., Mueser, M., Harstrick, A., Verslype, C., Chau, I. & Van Cutsem, E. (2004). Cetuximab monotherapy and cetuximab plus irinotecan in irinotecan-refractory metastatic colorectal cancer. *N Engl J Med* **351**, 337-45.
36. Bonner, J. A., Harari, P. M., Giralt, J., Azarnia, N., Shin, D. M., Cohen, R. B., Jones, C. U., Sur, R., Raben, D., Jassem, J., Ove, R., Kies, M. S., Baselga, J., Youssoufian, H., Amellal, N., Rowinsky, E. K. & Ang, K. K. (2006). Radiotherapy plus cetuximab for squamous-cell carcinoma of the head and neck. *N Engl J Med* **354**, 567-78.
37. Messersmith, W. & Hidalgo, M. (2007). Panitumumab, a Monoclonal Anti Epidermal Growth Factor Receptor Antibody in Colorectal Cancer: Another One or the One? *Clinical Cancer Research* **13**, 4664-4666.
38. Ackerman, M., Levary, D., Tobon, G., Hackel, B., Orcutt, K. D. & Wittrup, K. D. (2009). Highly avid magnetic bead capture: An efficient selection method for de novo protein engineering utilizing yeast surface display. *Biotechnol Prog*.

2. EVOLUTION THROUGH RECURSIVE MUTAGENESIS AND LOOP SHUFFLING

Introduction

Fn3 has been demonstrated as an effective scaffold for molecular recognition.[1-9] Modification of the amino acid sequence in these loops can impart novel binding capacity on the scaffold. Koide and colleagues have diversified the BC and FG loops and screened libraries by phage display to yield a 5 μM binder to ubiquitin[4] and a 250 nM binder to Src SH3 domain;[2] FG loop libraries yielded improved intregrin binders[8] and novel estrogen receptor binders,[3] the latter of which was identified by a yeast two-hybrid screen. Phage display and yeast surface display were used to screen a three-loop library with tyrosine/serine diversity to isolate binders of 5, 7, and 30 nM to yeast small ubiquitin-like modifier, human small ubiquitin-like modifier 4, and maltose-binding protein, respectively.[5] mRNA display was used to screen three-loop libraries to identify a 20 pM binder to tumor necrosis factor α[9] and a 340 pM binder to vascular endothelial growth factor receptor 2.[7] Lipovsek, *et al.* used yeast surface display to engineer a 350 pM binder to lysozyme from a library with diversified BC and FG loops.[6]

Though wild-type Fn3 is highly stable and monomeric, previously reported high affinity Fn3 clones are oligomeric or unstable. Two clones with 300 pM and 1 nM affinity for tumor necrosis factor α exhibited midpoints of proteolysis susceptibility at approximately 30° and 42°, respectively.[9] A clone with 340 pM affinity for vascular endothelial growth factor receptor 2 exhibited T_m values of 32-52°; a different clone was more stable but only 40-50% monomeric and was of moderate affinity (13 nM).[7] Stability engineering yielded a stable, monomeric 2 nM binder. The ability to engineer Fn3 domains with high affinity and stability represents a critical need in scaffold development since both properties are required for clinical and biotechnological applications.

Synthesis errors during our previous library creation resulted in rare clones with non-wild-type loop lengths.[6] These loop length variants were preferentially selected during binder isolation and binding titrations revealed their critical importance to binding (Figure 2.1). Moreover, sequence analysis reveals non-wild-type lengths in tumor necrosis factor α binders[9] and carcinoembryonic antigen binders (data not shown). The rarity of these length variants in the initial library resulted in very low coverage of theoretical length variant sequence space. These results suggest that more extensive sampling of length variant sequences could improve the binding capabilities of the Fn3 scaffold. This hypothesis is supported by the fact that CDR length diversity is already an acknowledged component of antibody engineering,[10; 11] and it has been demonstrated that antibody affinity maturation is improved with the inclusion of length diversity.[12; 13] Koide and colleagues incorporated loop length diversity in a recent Fn3 library and successfully obtained binders with length variation.[5] From a structural perspective, Batori *et al.* demonstrated that the BC, DE, and FG loops can tolerate insertions of four glycine residues while maintaining a native fold.[14] In addition, phylogenetic analysis reveals significant length diversity in each of the three loops across different fibronectin type III domains in multiple species (Figure 2.2). Thus, we hypothesize that loop length diversity will be stably tolerated and improve the functional binding capabilities of the Fn3 scaffold and perhaps could improve the stability/affinity relationship.

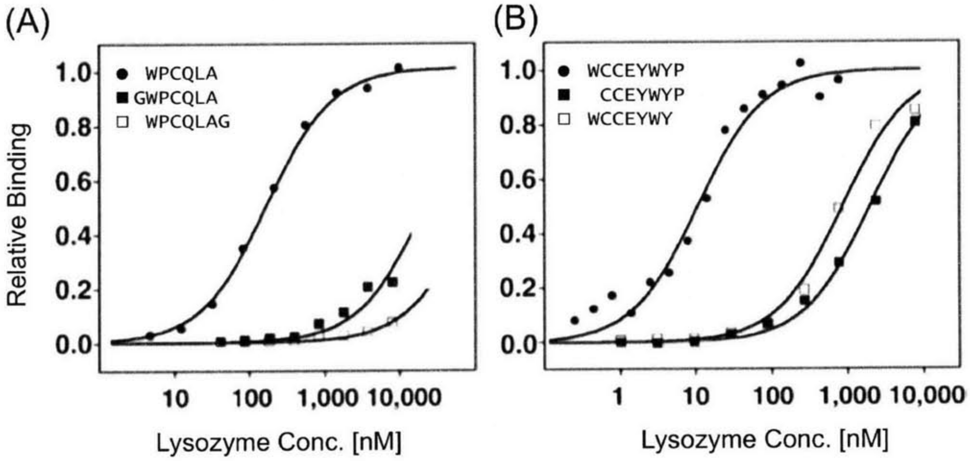


Figure 2.1. *Impact of loop length on affinity.* Biotinylated lysozyme-binding Fn3 domains of non-native loop length (circles) were compared to reversion mutants (squares) of wild-type length. Loop sequence is indicated.

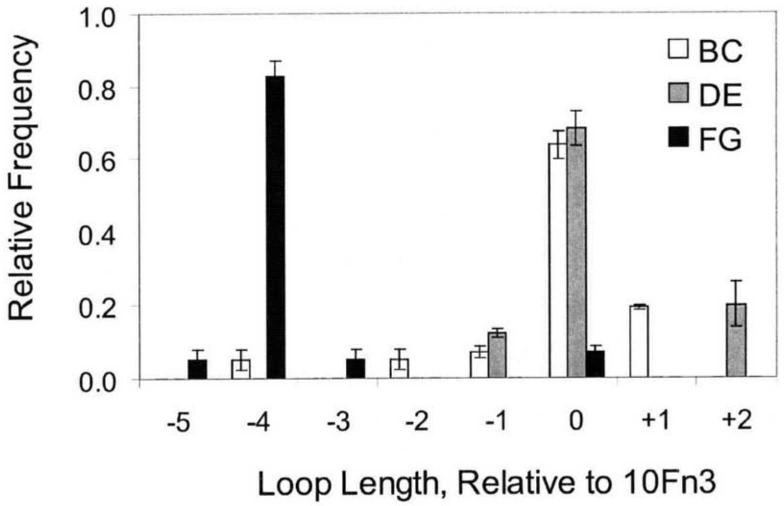


Figure 2.2. *Loop length variability of fibronectin type III domains.* The relative frequency of loop lengths of the BC (white), DE (gray), and FG (black) loops of sixteen type III domains of fibronectin, relative to the tenth type III domain (10Fn3) of human fibronectin, are shown. Values and error bars represent the average and standard deviation for five species: human, cow, rat, mouse, and chicken.

Yet loop length diversity increases theoretical sequence space, which is already immense, increasing the demand for efficient protein engineering methods both in terms of clonal selection and sequence diversification. Yeast surface display enables quantitative selections using fluorescence activated cell sorting (FACS) and correspondingly fine affinity discrimination.[15] The eukaryotic protein processing of yeast also improves the functional stringency of selections.[16] One potential limitation of yeast surface display is that cellular transformation efficiency constrains library size to 10^7-10^9 without excessive effort. *In vitro* display methods enable larger theoretical library size which correlates with selection of improved binders;[17; 18] however, as has been shown for phage display, nominal library size does not necessarily equate with functional diversity[16]. The intrinsic mutagenesis from the requisite PCR step in mRNA and ribosome display is a key contributor to the success of these display technologies.[19] We hypothesized that an increase in the frequency of mutagenesis during directed evolution would improve the efficiency of cellular display methods by increasing the breadth of the sequence space search in the vicinity of many lead clones, rather than only a select few.

Another important engineering component is the manner in which selected sequences are diversified throughout directed evolution. Successful techniques include DNA shuffling,[20] CDR shuffling,[21; 22] CDR walking,[23] and error-prone PCR mutagenesis[24] either towards the entire gene or directed to suspected paratopes. In the current work, we combine error-prone PCR and an analog to CDR shuffling, both directed at the expected paratope and throughout the Fn3 domain, to yield both mild and significant changes in sequence space.

We demonstrate that loop length diversity and a novel affinity maturation scheme enable robust and efficient selection of stable, high affinity binders to lysozyme. The binders are characterized in terms of binding, stability, and structure including detailed analysis of the molecular basis of binding of a picomolar affinity clone.

Results

Fn3 library construction

A library was created in which eight, five, and ten amino acids of the BC, DE, and FG loops, respectively, were diversified both in amino acid length (Figure 2.2) and composition (Figure 2.3). Amino acid composition was randomized using NNB degenerate codons to yield all 20 amino acids with reduced stop codon frequency. Four different loop lengths were chosen for each loop based partially on the loop lengths observed in fibronectin type III domains in multiple species (Figure 2.2). The library of Fn3 genes was incorporated into a yeast surface display system by homologous recombination with a vector incorporating an N-terminal Aga2 protein for display on the yeast surface and a C-terminal c-myc epitope for detection of full length Fn3.[25] Library transformation yielded 6.5×10^7 yeast transformants. Sixteen of 26 clones sequenced (62%) matched library design. Nine (35%) contained frameshifts and one (4%) was annealed improperly or underwent unintentional homologous recombination in yeast. NNB diversification of the loops yields stop codons in approximately 44% of clones. Thus, 34% (16/26 x (1-0.44)) of transformed cells should display full length Fn3. This percentage was verified by flow cytometry analysis (data not shown). The library contains approximately 2.3×10^7 (6.5×10^7 x 0.34) full-length Fn3 clones.

Loop	Residues	Wild-type Sequence	Lengths*
BC	23-30	DAPAVTVR	-2, -1, 0, +1
DE	52-56	GSKST	-1, 0, +1, +2
FG	77-86	GRGDSPASSK	-5, -4, -2, 0

* amino acid length relative to wild-type

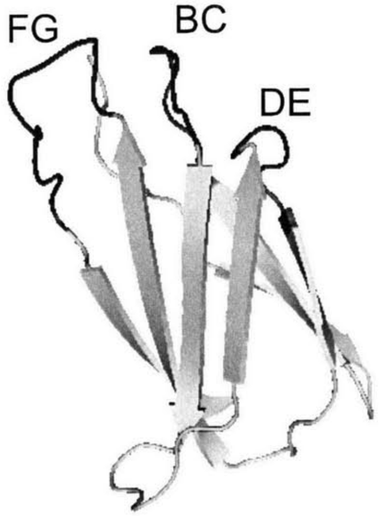


Figure 2.3. *Library design.* The naïve library is randomized in the BC, DE, and FG loops (structure schematic derived from Main, *et al.*[31]) at the indicated positions to four possible loop lengths each.

Population mutagenesis design

A negative side effect of loop length diversity is a further increase in the vastness of possible protein sequence space to 10^{34} possible amino acid sequences for the three loops. Thus, the 2.3×10^{7} clones in the original library grossly undersample sequence space. Therefore, an extensive diversification method was desired to broadly search sequence space during affinity maturation. Error-prone PCR directed solely at the solvent-exposed loops was used to focus diversity on the likely paratope. Additionally, to make substantial changes in sequence, a loop shuffling approach was developed. Fn3 genes were constructed with conserved wild-type framework sequence and randomly shuffled mutated loops from the pool of selected clones. Yet, as effective clones are selected, such substantial sequence modification is not desired. Thus, error-prone PCR without shuffling was also employed. This mutagenesis targeted the entire gene to also enable selection of beneficial framework mutations. Both mutagenesis strategies were used in parallel at each mutagenesis step throughout affinity maturation (Figure 2.4). Analysis of selected clones after the fact indicates that both loop shuffling and whole-gene error-prone PCR contributed to improved phenotypes.

The Fn3 library was screened using yeast surface display and FACS. Plasmids from thousands of lead clones in a partially screened yeast population were collected by yeast zymoprep. Quantitative real-time PCR revealed that full population diversity was recovered (data not shown). Both error-prone PCR of the full gene and shuffling of mutagenized loops were successfully employed to diversify the population of lead clones (Figure 2.4B). Gene mutagenesis was performed by error-prone PCR with nucleotide analogs.[26] Independent error-prone PCRs were conducted for each of the three loops using primers that overlap either the adjacent loop primer or the plasmid vector. This overlap enabled shuffled gene reconstruction via homologous recombination during yeast transformation (Figure 2.4C). The number of diversified transformants during affinity maturation ranged from 2 to 20 million (mean 8.1 million) for full gene mutagenesis and 0.1 to 6.5 million (mean 2.5 million) for loop mutagenesis. Sequence analysis exhibited mutagenesis that matched the desired 1 to 5 amino acid mutations per gene (data not shown), which was achieved with a single combination of nucleotide analog concentration and number of PCR cycles determined by mathematical modeling (Appendix A).

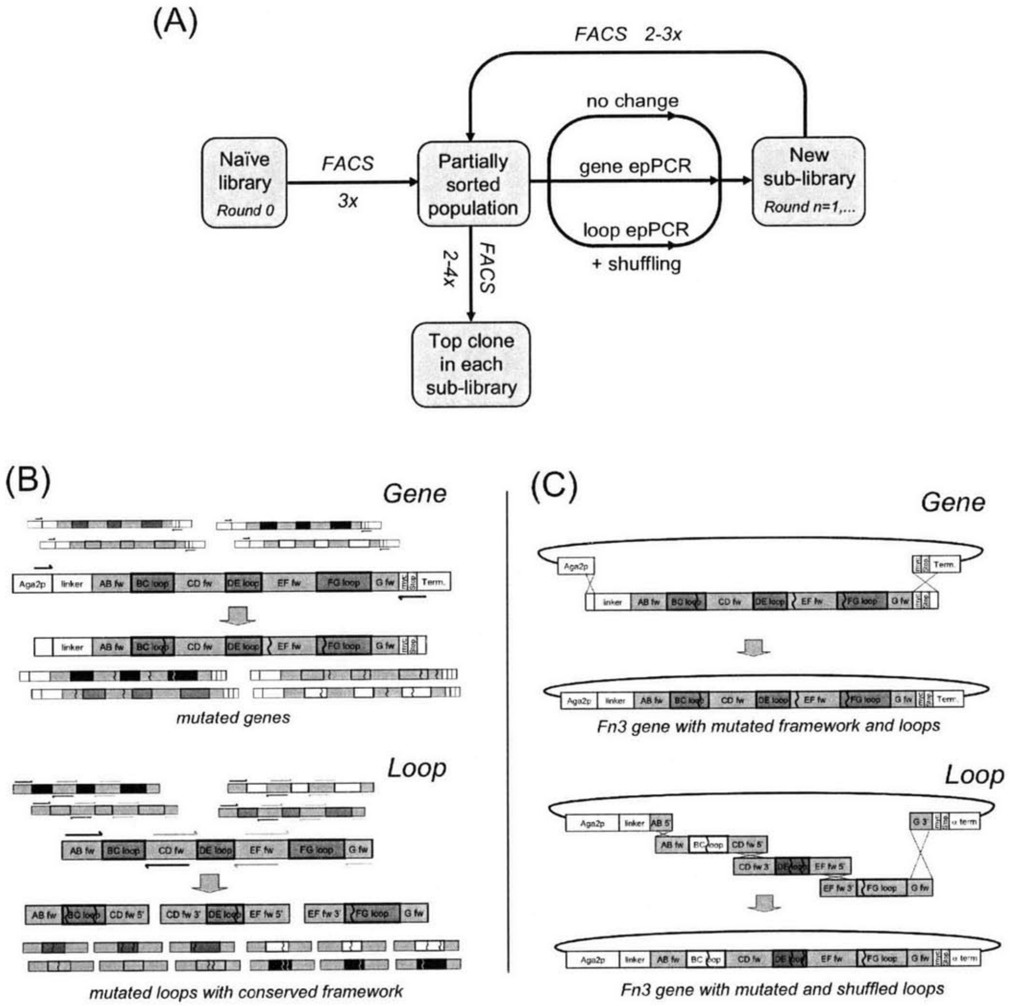

(A)
FACS 2-3x
no change
Naïve library
Round 0
FACS
3x
Partially sorted population
gene epPCR
loop epPCR
+ shuffling
New sub-library
Round n=1,...
2-4x
FACS
Top clone in each sub-library
(B)
Gene
Aga2p
linker
AB fw
BC loop
CD fw
DE loop
EF fw
FG loop
G fw
Term.
mutated genes
Loop
mutated loops with conserved framework
(C)
Gene
Fn3 gene with mutated framework and loops
Loop
Fn3 gene with mutated and shuffled loops

Figure 2.4. *Affinity maturation scheme.* (A) The naïve library is sorted three times by FACS for binding to lysozyme. The partially sorted population (*i.e.,* the population is still diverse) is mutated by error-prone PCR (epPCR) of the entire gene or of the loops alone, which are then shuffled during gene formation by homologous recombination. These two mutant populations are combined with the unmutated population to yield a new sub-library. The process of FACS and mutagenesis is then cyclically repeated. The rounds of affinity maturation are named for the number of mutagenesis cycles they have faced; thus sorting of the naïve library is round 0, followed by rounds 1, 2, ... A typical round of isolation and mutagenesis requires about one week although kinetic sorting as well as equilibration at picomolar concentrations lengthens the duration. Rounds 0-3 were completed in two months and rounds 4-8 were completed in two additional months. The highest affinity clone present in each round of maturation is identified by further sorting and sequence analysis. (B) Error-prone PCR is used to introduce random mutations either into the entire Fn3 gene or the three loops (separate PCRs). Arrowed lines indicate PCR primers; matching shades correspond to primers in a single PCR reaction. The framework (fw) and loop regions of the gene are indicated. Zigzag lines represent nucleotide mutations. Smaller unlabeled images are used to emphasize that PCR templates and products are part of a repertoire based on many gene variants. (C) Two yeast sub-libraries are created by transformation via electroporation. During transformation, the full plasmid is created by homologous recombination of the linearized vector with either the single mutated gene insert or the three mutated loop inserts.

High affinity binder engineering

The efficacy of the new library and affinity maturation scheme was tested by engineering binders to lysozyme, a model protein that we have previously used as an Fn3 target with different library and maturation approaches.[6] The naïve library was sorted three times by FACS using multivalent biotinylated lysozyme preloaded on streptavidin-fluorophore conjugates. The resultant population was diversified by both gene and loop mutagenesis. Transformed mutants as well as the original clones were sorted twice by FACS yielding enrichment of evident lysozyme-binding clones (Figure 2.5). Four additional rounds of isolation and maturation, each consisting of 2-3 FACS selections followed by mutagenesis (Figure 2.4), were conducted using monovalent lysozyme, ranging from 1 μM to 20 pM lysozyme. Labeling at low picomolar concentrations while maintaining stoichiometric excess of target to Fn3 would require impractically large volumes. Therefore, kinetic competition sorting was used for the final three rounds of isolation and maturation. The final two sub-libraries were sorted an additional three times and the collected populations were sequenced to identify the highest affinity clones.

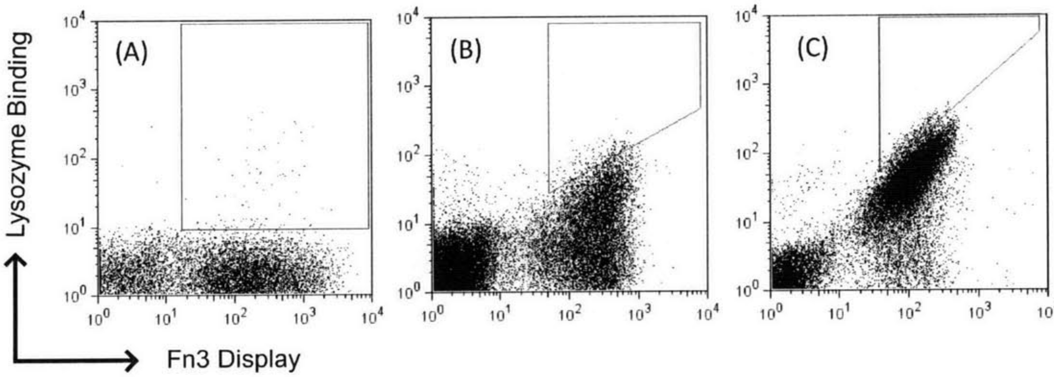


Figure 2.5. *Binder isolation and affinity maturation.* Yeast libraries displaying Fn3 were labeled with mouse anti-c-myc antibody followed by goat anti-mouse fluorophore as well as biotinylated lysozyme and streptavidin-fluorophore and analyzed by flow cytometry. (A) Yeast population during second round of isolation and maturation labeled with 50 nM multivalent lysozyme preloaded in a 3:1 stoichiometry on streptavidin-R-phycoerythrin. (B) Yeast population during sixth round of isolation and maturation labeled with 0.5 nM monovalent lysozyme followed by streptavidin-AlexaFluor488. (C) Yeast population during eighth round of isolation and maturation labeled with 2 nM monovalent lysozyme for 15 min. followed by 35 h of dissociation and labeling with streptavidin-R-phycoerythrin. Polygonal regions represent sort gates for cell selection.

Three dominant clones were identified by sequence analysis (Table 2.1). Affinity titrations indicate that all three clones have comparable equilibrium dissociation constants (K_d) for binding to biotinylated lysozyme: 2.6 ± 0.6 pM (clone L7.5.1), 2.9 ± 1.4 pM (L8.5.2), and 2.8 ± 0.5 pM (L8.5.7) (Figure 2.6A, Table 2.1). The high affinity binding was validated using purified Fn3 domains in an equilibrium competition titration, which indicated an affinity of 6.9 ± 0.3 pM for L7.5.1 (Figure 2.6B). Dissociation rates for the three clones range from 2.5-5.4 $\times 10^{-6}$ s^{-1}, which correspond to dissociation half times of 36-78 h (Figure 2.6C). The association rate of L7.5.1 was measured experimentally as 2.0 ± 0.5 $\times 10^{6}$ $M^{-1}s^{-1}$ (Figure 2.6D). The calculated value of k_{off} / k_{on} is 1.2 ± 0.3 pM, which is reasonably consistent with the experimentally measured K_d of 2.6 ± 0.6 pM. In addition to high affinity, the clones exhibit target specificity as they do not show appreciable binding to an array of other molecules (Figure 2.7).

Table 2.1. Characterization of wild-type, L7.5.1, L8.5.2, and L8.5.7.

Clone	Amino Acid Sequences				K_d [pM]	k_{off} [10^{-5} s^{-1}]	Monomer	T_m [ºC]		$T_{1/2}$ [ºC]
	BC Loop	DE Loop	FG Loop	Framework				DSC	CD	
WT	DAPAVTVR	GSKST	GRGDSPASSK	-	>10^7	n/d	99%	85.7	84.2	n/d
L7.5.1	RGYPWAT	GVTN	RVGRTFDTPG	P15S, R33G, T35I, P44L, V50M	2.6 ± 0.6	0.25 ± 0.02	99%	58.1	58.8 ± 1.6	55.7 ± 0.1
L8.5.2	RGCPWAI	GVTN	RVGRMLCAPG	R33G, I34V, N42S, P44L, V45A, V50M, K63E, K98R	2.9 ± 1.4	0.32 ± 0.01	80%	n/d	52.5 ± 0.2	50.6 ± 0.5
L8.5.7	RDRPWAI	GVTN	RLSIVPYA	D3G, L18I, R33G, N42S, P44L, V50M, Y73H, N91T, S100P	2.8 ± 0.5	0.54 ± 0.06	93%	54.5	n/d	53.0 ± 0.4

K_d: equilibrium dissociation constant at 25°

k_{off}: dissociation rate constant at 25°

Monomer: percent of purified protein present as monomer in analytical size exclusion chromatography

T_m: midpoint of thermal denaturation curve as determined by differential scanning calorimetry (DSC) or circular dichroism (CD)

$T_{1/2}$: midpoint of thermal denaturation curve as determined by yeast surface display residual activity assay

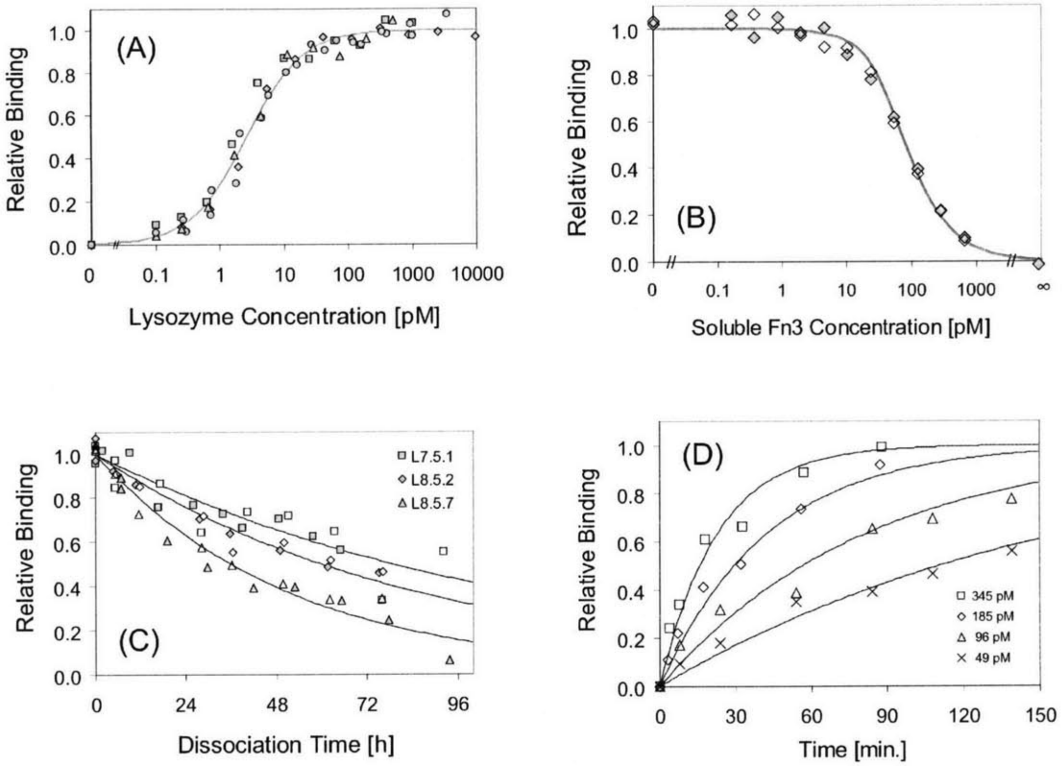


Figure 2.6. *Determination of binding parameters for high affinity clones.* L7.5.1, L8.5.2, or L8.5.7 was displayed on the yeast surface and assayed for binding to biotinylated lysozyme at 25°. Different symbols indicate replicate experiments. Solid lines indicate theoretical values. (A) Fraction of displayed L7.5.1 binding to biotinylated lysozyme at equilibrium. Only L7.5.1 is shown for simplicity. (B) Relative binding of 20 pM biotinylated lysozyme to displayed L7.5.1 at equilibrium in the presence of indicated amount of soluble L7.5.1 competitor. ∞ indicates samples without biotinylated lysozyme. (C) Fraction of displayed L7.5.1 (squares), L8.5.2 (diamonds), or L8.5.7 (triangles) binding to biotinylated lysozyme after dissociation at 25° for the indicated time. (D) Fraction of displayed L7.5.1 binding to biotinylated lysozyme after association at 25°; biotinylated lysozyme was present at 49 pM (crosses), 96 M (triangles), 185 pM (diamonds), or 345 pM (squares). Results are presented from a single experiment, which is representative of triplicate experiments performed at different sets of concentrations.

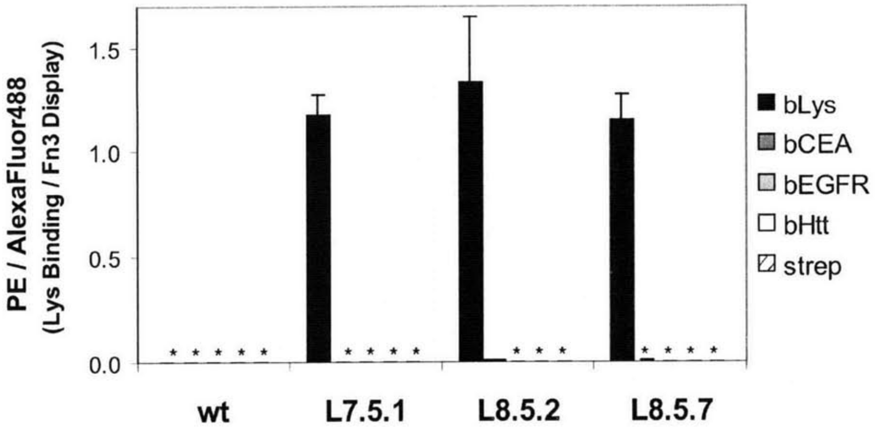


Figure 2.7. *Binding specificity.* Wild-type Fn3 (wt), L7.5.1, L8.5.2, or L8.5.7 was displayed on the yeast surface, washed, and incubated with 100 pM biotinylated lysozyme (black) or 100 nM of non-target molecules: biotinylated carcinoembryonic antigen (bCEA, dark gray), biotinylated EGFR domain IV (bEGFR, light gray), biotinylated huntingtin peptide (bHtt, white), or streptavidin-R-phycoerythrin (strep, striped). Binding was detected with streptavidin-R-phycoerythrin secondary labeling and flow cytometry. The data are presented as the mean fluorescence units from binding (PE signal) normalized by Fn3 display (AlexaFluor 488 signal). Values and error bars represent mean ± standard deviation of duplicate measurements. * indicates a value less than 0.005.

Since previously reported high affinity clones were unstable or oligomeric, it was desired to examine the behavior of the current Fn3 domains. The clones were produced in bacterial culture and purified for biophysical characterization. Analytical size exclusion chromatography indicated that all three clones are predominantly monomeric with only L8.5.2 present in a significant oligomeric state (80% monomeric, Table 2.1, Figure 2.8A). L8.5.2 contains two cysteine residues and intermolecular disulfide bonding contributes to oligomerization as indicated by non-reducing SDS-PAGE (data not shown). In addition, far-UV circular dichroism analysis reveals no significant differences in secondary structure between wild-type Fn3 and L7.5.1 indicating that despite loop mutation and length variation the structural integrity of the domain is maintained (Figure 2.8B). The thermal stabilities of the proteins were analyzed by differential scanning calorimetry (DSC) and circular dichroism of purified protein as well as thermal denaturation of protein displayed on the yeast surface (Table 2.1, Figure 2.9). L7.5.1 is the most stable clone with midpoints of thermal denaturation ranging from 55.7-58.8° for the three methods. Denaturation of L7.5.1 is not reversible because of aggregation at high temperatures. Both other clones are also stable with T_m values over 50°.

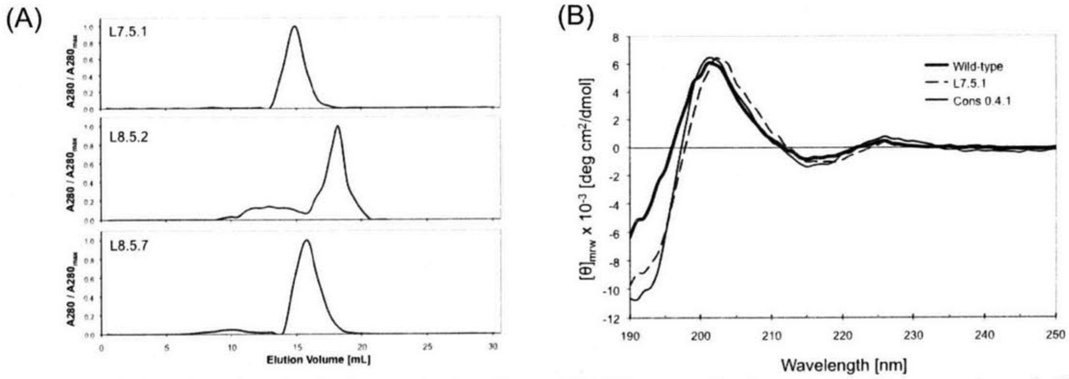


Figure 2.8. *Biophysical characterization.* (A) Size exclusion chromatography elution profiles. Clones L7.5.1, L8.5.2, and L8.5.7 were analyzed on a Superdex 75 HR10/300 column. The absorbance at 280 nm of the elution is normalized for each clone. (B) Far-UV circular dichroism. Ellipticity was measured from 250 to 190 nm for 8-10 μM solutions of wild-type (thick solid line), L7.5.1 (dashed line), or Cons0.4.1 (thin solid line). Spectra are presented as mean residue ellipticity (θ_{mrw}), which is obtained by factoring in molecular mass, concentration, and 1 mm path length.

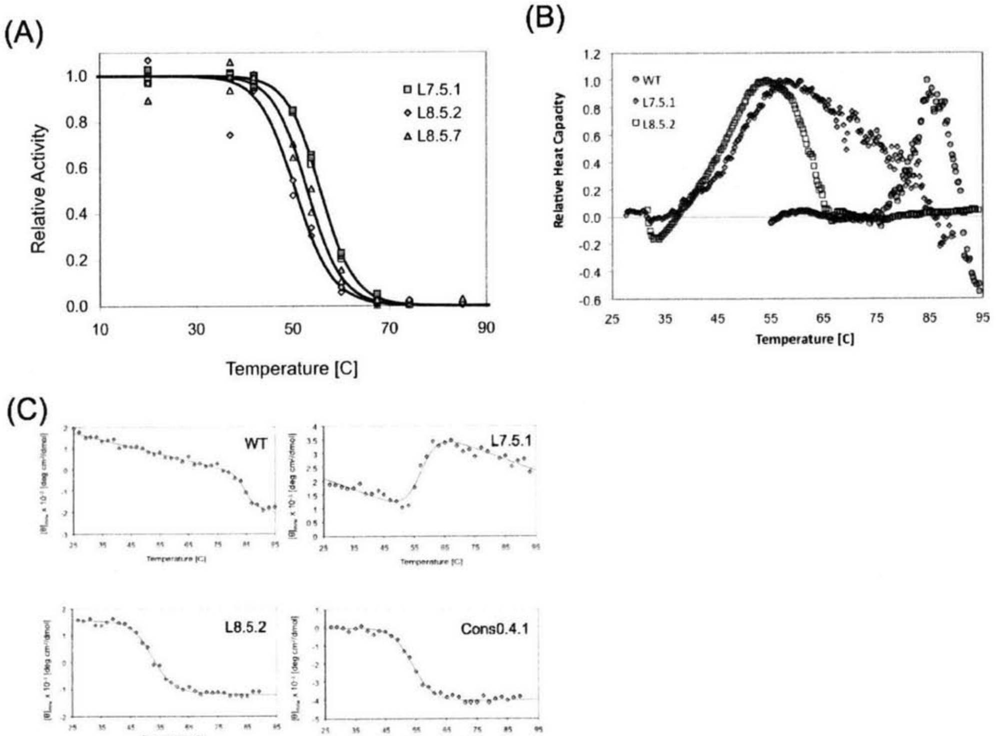


Figure 2.9. *Thermal denaturation.* (A) Thermal denaturation on the yeast cell surface. Yeast displaying L7.5.1 (squares), L8.5.2 (diamonds), or L8.5.7 (triangles) were incubated at the indicated temperature for 30 min. followed by a flow cytometric assay for biotinylated lysozyme binding. (B) Differential scanning calorimetry. Protein was heated from 25° to 95° at a rate of 1° per minute. The buffer scan is subtracted from the protein scan; heat capacities are normalized for each clone. Irreversible aggregation of L7.5.1 occurs at high temperatures. (C) Circular dichroism. The ellipticity at 216 nm was measured from 25° to 95°. Diamonds represent experimental data. Solid lines represent a two-state unfolding curve.

Analysis of intermediate populations

Intermediate populations were investigated to elucidate the progress of affinity maturation. Each sub-library (*i.e.*, library of mutagenized clones created during affinity maturation) was sorted by FACS without mutagenesis to identify the highest affinity clone at each stage of affinity maturation. The highest affinity clone was identified as the dominant clone by sequencing several clones from the extensively sorted sub-libraries. Sequences, equilibrium dissociation constants and dissociation rate constants were determined (Table 2.2). The highest affinity binder in the original library exhibits apparent mid-micromolar affinity. However, affinity maturation progressively improves binding performance to yield nanomolar and then picomolar binders (Figure 2.10). It is noteworthy that affinity maturation exhibits a relatively consistent correlation to the number of clones analyzed throughout the course of affinity maturation until the final round of directed evolution. A steady increase characteristic of a consistently progressing affinity maturation scheme is observed rather than one single exceptional increase in affinity indicative of a fortuitous mutation or recombination.

Table 2.2. Highest affinity Fn3 domains in each sub-library.

Round	Amino Acid Sequences				K_d [pM]	k_{off} [10^{-5} s^{-1}]	$t_{1/2}$ [h]
	BC Loop	DE Loop	FG Loop	Framework			
0	RDCPWAT	WTPVCF	SSQRGCM	none	>>100,000	n/d	n/d
1	SLDNQAN	GQSD	RCEPSRNSAV	none	>100,000	n/d	n/d
2				same clone as round 1			
3	SLDNQAN	GVTN	RVGRMLDTPG	P44S, V50M	7600 ± 1100	460	0.04 h
4	SLDNQAK	GATN	RCKPFRNSAV	P44S, V50M, T97I	330 ± 50	n/d	n/d
5	RDCPWAI	GVTN	RVGRMSCTSG	V1A, S1P, T14A, R33G, P44L, V50M	16 ± 6	4.5 ± 0.3	4.2 ± 0.3 h
6	RGCPWAI	GVTN	RVGRMLCTPG	P15S, R33G, N42S, P44L, V50M, K63E	6.6 ± 1.3	0.72 ± 0.07	27 ± 3 h
7	RGYPWAT	GVTN	RVGRTFDTPG	P15S, R33G, T35I, P44L, V50M	2.6 ± 0.6	0.25 ± 0.02	78 ± 1 h
8 (A)	RGCPWAI	GVTN	RVGRMLCAPG	R33G, I34V, N42S, P44L, V45A, V50M, K63E	2.9 ± 1.4	0.32 ± 0.01	60 ± 1 h
8 (B)	RDRPWAI	GVTN	RLSIVPYA	D3G, L18I, R33G, N42S, P44L, V50M, Y73H, N91T	2.8 ± 0.5	0.54 ± 0.06	36 ± 4 h

Round: number of mutagenesis cycles (round 0 indicates naïve library)

K_d: equilibrium dissociation constant at 25°

k_{off}: dissociation rate constant at 25°

$t_{1/2}$: time for 50% dissociation of Fn3:lysozyme complex at 25°

n/d: not determined

^{a,b}: Two clones with similar affinities were identified from round 8

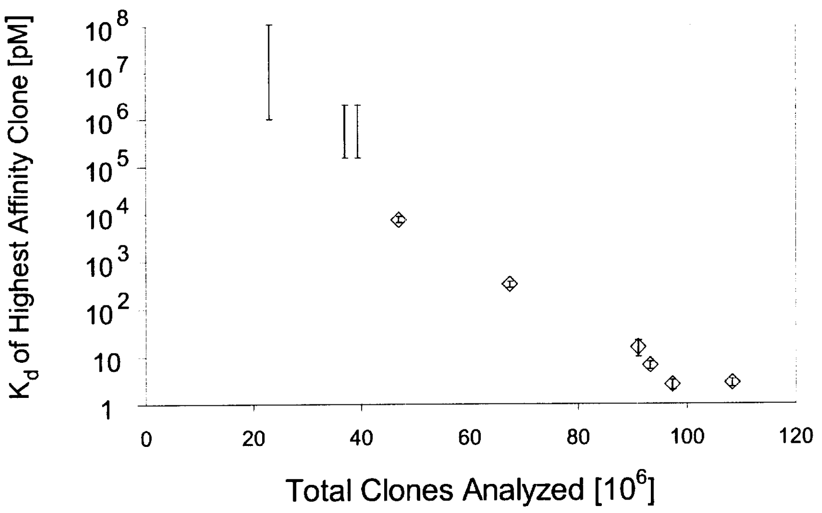


Figure 2.10. *Affinity maturation progress.* The highest affinity clone in each affinity maturation sub-library was identified by FACS. The equilibrium dissociation constant of the first three clones could not be determined accurately by yeast surface display titration because of the relatively low affinity. Thus, an estimated range of possible affinities consistent with equilibrium labeling at high nanomolar concentrations is indicated. The equilibrium dissociation constant of the latter six clones was determined by titration and is represented as the mean ± one standard deviation of replicate measurements. The number of total clones analyzed is the cumulative total of the number of full-length clones in the naïve library and total yeast transformants in all affinity maturation sub-libraries.

The impact of loop shuffling is evident in multiple cases. The BC loop from the round 1 clone is recombined with new DE and FG loops in round 3 to yield an 8 nM binder. A mutated version of the FG loop present in round 1 is then recombined with mutated BC and DE loops from round 3 to achieve picomolar binding. The highest affinity clones in rounds 5-8 result from apparent shuffling of the BC loop observed in round 0 and the DE and FG loops observed in round 3 as well as multiple framework mutations. The appearance of point mutations is also apparent, both within the loops and in the framework region. The impact of these framework mutations was investigated in more detail in the context of L7.5.1.

L7.5.1 analysis

L7.5.1 was selected for more thorough analysis because it has the fewest framework mutations and is the most stable of the high affinity clones. The engineered elements of the clone were analyzed for their impact on affinity and stability. Each randomized loop and framework mutation was independently restored to wild-type sequence and both affinity and stability were measured using yeast surface display. Reversion of the FG loop had the strongest effect on binding as wild-type restoration decreased affinity to the micromolar level. Conversely, DE loop reversion had a nearly negligible effect on binding, and the BC loop reversion had an intermediate effect with a 4700-fold reduction (Figure 2.11A). Both BC and DE loop reversion significantly destabilize the domain (Figure 2.11B). Though these loop sequences confer high thermal stability in the wild-type context, modification of adjacent loops apparently adjusts the intramolecular contacts. Thus, during multi-loop diversification, selected sequences must provide not only proper intermolecular contacts for high affinity binding but also proper intramolecular contacts for structural integrity and stability. The stability of the FG loop reversion could not be accurately determined by this method because of its weak binding.

Three of the five framework mutations in the selected clone (T35I, P44L, and V50M) are relatively conservative mutations and were beneficial to both affinity and stability (Table 2.3). Conversely, R33G replaces a large, positively charged side chain with a single hydrogen; this mutation provides 28-fold stronger binding without a substantial effect on stability. Perhaps the arginine side chain was not accommodated sterically and/or electrostatically at the Fn3:lysozyme interface or the glycine enables a beneficial backbone conformation. The P15S mutation had only a very minor affinity improvement but substantially destabilized the domain. The lack of impact on binding is reasonable given its distance from the expected paratope, and the substantial destabilization is not surprising given the removal of the backbone-constraining proline in the loop between the A and B strands. Selection of this mutation was likely coincident with another beneficial mutation as it does not impart significant benefit. It should be noted that while the most beneficial mutations, R33G, P44L, and V50M, are consistently observed throughout maturation (Table 2.2), P15S and T35I are rare. Overall, these data suggest the paratope is focused on the BC and FG loops as well as potential key contacts on the C and D strands of the β sheet.

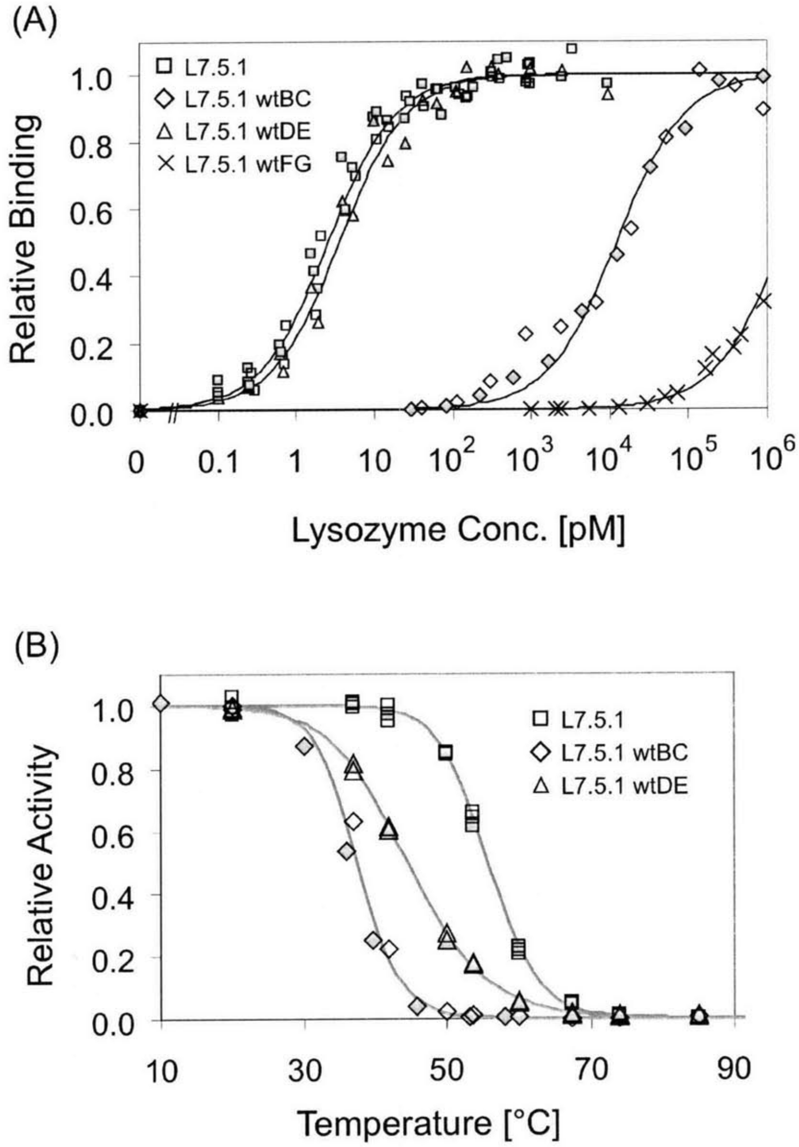


Figure 2.11. *L7.5.1 loop analysis.* Each engineered loop of L7.5.1 was independently restored to wild-type sequence and the resultant clones were displayed on the yeast surface. Filled and empty symbols indicate replicate experiments. (A) Binding to biotinylated lysozyme at 25° was quantified by flow cytometry. Clone and equilibrium dissociation constant: L7.5.1 (squares): 2.6 ± 0.6 pM; L7.5.1wtBC (diamonds): 12.4 ± 0.7 nM; L7.5.1wtDE (triangles): 3.7 ± 1.4 pM; L7.5.1wtFG (crosses): 1.6 ± 0.2 μM. (B) Yeast displaying Fn3 were incubated at the indicated temperature for 30 min. followed by a flow cytometric assay for biotinylated lysozyme binding. Clone and midpoint of thermal denaturation: L7.5.1 (squares): 55.7 ± 0.1°; L7.5.1wtBC (diamonds): 37.4 ± 1.5°; L7.5.1wtDE (triangles): 44.5 ± 0.1°.

Table 2.3. Affinity and stability of framework mutation reversions of L7.5.1. The five framework mutations of L7.5.1 were individually reverted to wild-type amino acids. Binding affinity and thermal stability were determined by yeast surface display.

AA	WT	L7.5.1	K_d [pM]	$T_{1/2}$ [°C]	Location
L7.5.1 - Parent Clone			2.6 ± 0.6	55.7 ± 0.1	-
15	Pro	Ser	3.4 ± 0.1	59.5 ± 0.7	AB loop
33	Arg	Gly	74 ± 23	56.2 ± 0.6	C strand
35	Thr	Ile	5.8 ± 1.6	54.5 ± 0.5	C strand
44	Pro	Leu	14.3 ± 4.0	50.3 ± 1.2	CD loop
50	Val	Met	25 ± 16	53.7 ± 1.8	D strand

AA: amino acid position
WT: wild-type side chain
L7.5.1: clone 7.5.1 side chain
K_d: equilibrium dissociation constant at 25°
$T_{1/2}$: midpoint of thermal denaturation as determined by yeast surface display

Focused mutagenesis

The conservation of the DE loop sequence throughout much of the affinity maturation (Table 2.2) despite its relatively low impact on affinity relative to wild-type prompted further study of this loop. The functionally tolerable diversity and the potential for improved binding were explored through re-randomization of the loop sequence. Within the context of L7.5.1 S15P, the loop was randomized using NNB nucleotide diversity and the same four loop length options as the original library. Labeling the library with 30 pM lysozyme, which yields 90% binding to the parent clone L7.5.1 S15P, yields many clones that provide effective binding indicating that the loop can tolerate significant diversity while maintaining binding. Specifically, 7% of clones exhibit binding at 75% of maximum, which is characteristic of a 10 pM K_d, or within three-fold of the parent clone (Figure 2.12). Yet, 30% of the full-length clones bind at less than 1% of maximum indicating that many loop sequences can greatly hinder binding either through direct interaction with the binding partner or through structural modification of the Fn3 paratope.

In addition, the library was sorted for binding to biotinylated lysozyme using FACS. Sequence analysis after four selections identified a single clone, DE0.4.1, with GDLSHR replacing GVTN in the DE loop. Binding and stability analyses indicate a 3.1x improvement in binding to 1.1 pM at the expense of an 8.5° decrease in the $T_{1/2}$ (Table 2.4). The maintenance of glycine at position 52 from wild-type to L7.5.1 to DE0.4.1 is noteworthy for possible future library design especially considering the adjacent amino acid is proline. The proline-glycine pair provides an effective turn motif to begin the DE loop.

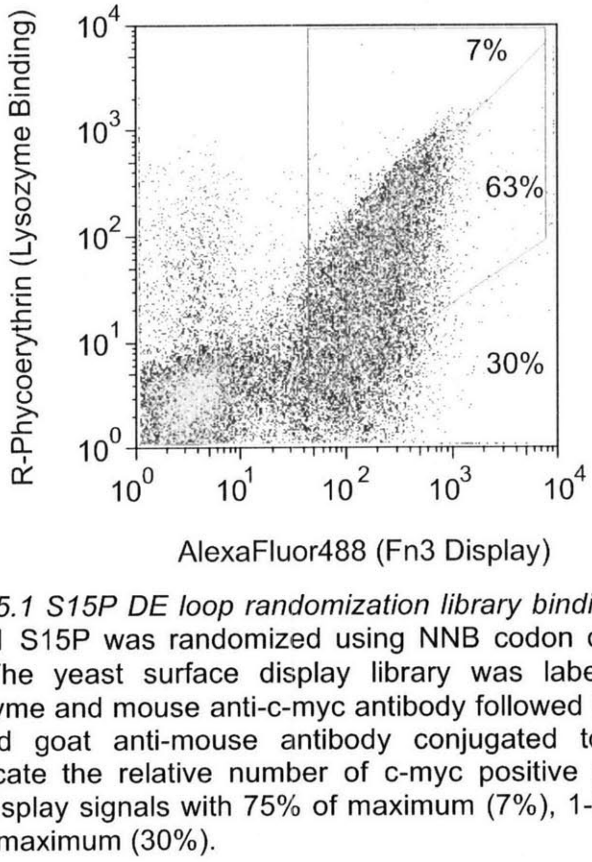


Figure 2.12. *L7.5.1 S15P DE loop randomization library binding analysis.* The DE loop of L7.5.1 S15P was randomized using NNB codon diversity with four loop lengths. The yeast surface display library was labeled with 30 pM biotinylated lysozyme and mouse anti-c-myc antibody followed by streptavidin-R-phycoerythrin and goat anti-mouse antibody conjugated to AlexaFluor488. Percentages indicate the relative number of c-myc positive clones with bLys binding : c-myc display signals with 75% of maximum (7%), 1-75% of maximum (63%), or <1% of maximum (30%).

Table 2.4. Affinity and stability of clones from focused mutagenesis.

Clone	Amino Acid Sequences				K_d [pM]	$T_{1/2}$ [°C]
	BC Loop	DE Loop	FG Loop	Framework		
L7.5.1 S15P	RGYPWAT	GVTN	RVGRTFDTPG	R33G, T35I, P44L, V50M	3.4 ± 0.1	59.5 ± 0.7
Cons 0.4.1	REDPWAK	GVTN	RVGWASYTLG	R33G, T35I, P44L, V50M	1.1 ± 0.6	59 ± 3
DE 0.4.1	RGYPWAT	GDLSHR	RVGRTFDTPG	R33G, T35I, P44L, V50M	1.1 ± 0.5	51 ± 3

K_d: equilibrium dissociation constant at 25°

$T_{1/2}$: midpoint of thermal denaturation curve as determined by yeast surface display residual activity assay

Lack of improvement from round 7 to round 8 in conjunction with sequence similarities during affinity maturation prompted investigation into a complementary method of affinity maturation. Since only a fraction of protein sequence space is accessible through single nucleotide mutations, a library was constructed in which the DNA encoding non-conserved amino acids in the BC and FG loops was randomized by degenerate oligonucleotides. Non-conserved amino acids were identified as the positions of L7.5.1-homologous clones at which sequence diversity was observed in any sequence during affinity maturation or sub-library analysis. The non-conserved amino acids were randomized in two modes: either to all 20 amino acids using NNB codons or to amino acids observed during sequence analysis using tailored codons at each position. The library was sorted for binding to biotinylated lysozyme using FACS. Sequence analysis after four selections yielded a single clone, Cons0.4.1. The affinity of Cons0.4.1 was measured as 1.1 ± 0.6 pM by yeast surface display titration and 0.33 ± 0.15 pM by equilibrium competition with purified Fn3 domain. The midpoint of thermal denaturation was measured as $52.5 \pm 2.1^{\circ}$ by circular dichroism analysis and $58.8 \pm 3.4^{\circ}$ by yeast surface display thermal resistance assay. Far-UV circular dichroism indicates that Cons0.4.1 maintains a secondary structure similar to that of the wild-type Fn3 domain (Figure 2.8B). Two of the eight codon changes were not possible with a single nucleotide mutation making this clone unlikely to be reached by error-prone PCR. Thus, while error-prone PCR provides a highly effective means of diversification, an improvement was observed through more thorough searching of a focused region of sequence space once a consistent binding motif was identified.

Discussion

In this work we explore the impact of multiple components of engineering a high affinity binding site in the Fn3 scaffold. The combination of loop length diversity and recursive mutagenesis including mutated loop shuffling enabled selection of the highest affinity Fn3 domains yet reported despite a relatively modest initial library size of 2.3×10^7 full-length Fn3 clones. The results extend the affinity attainable by this single domain scaffold further validating its use as a molecular recognition scaffold. In addition, insights were gained for both Fn3 design and engineering and protein engineering in general.

Analysis of the highest affinity binders in each sub-library (Table 2.2) demonstrates deviation from wild-type loop lengths in all three loops implicating the importance of this diversity element in Fn3 library design. Interestingly, the observed BC loops are all one amino acid shorter than wild-type, which is observed both in other fibronectin type III domains (Figure 2.1) and in a previously engineered binder.[6] DE loops are observed with either one less or one more amino acid than wild-type. Clone DE0.4.1 and the highest affinity clone from the naïve library (Table 2.2) justify inclusion of a six-amino acid DE loop despite its lack of existence in native fibronectin type III domains. FG loops of wild-type length as well as both two- and three-amino acid reductions were observed. The observed length diversity, as well as the success of length diversity in a restricted diversity Fn3 library published during this thesis,[5] support the inclusion of length diversity in all future Fn3 engineering. Moreover, the combination of phylogenetic analysis and sequence analysis of engineered binders should continue to elucidate the relative preference of each length to further improve library design.

Loop length diversity increases sequence space to 10^{34} possible amino acid sequences for the three loops though only 2.3×10^7 clones are present in the yeast library. The two modes of diversity introduced during recursive mutagenesis effectively search this large sequence space despite the sparse sampling. Sequence and binding analyses indicate that each of the elements of the mutagenesis method were beneficial. The loop mutagenesis path of the diversification was effective as both loop shuffling and loop-focused point mutations are evident (Table 2.2). In addition, the gene mutagenesis path was advantageous as an effective loop combination was maintained in rounds 5-8 and beneficial framework mutations were introduced throughout (Tables 2.2 and 2.3). The importance of diversifying many clones during each round of maturation is evident since homologs of the eventual DE and FG loops were not present in 30 sequences from the enriched populations from the naïve library. As a result of these combined components, affinity maturation rapidly and efficiently progressed to yield the clones with 3 pM affinity without rational intervention. The affinity maturation procedure is straightforward and simple; plasmid recovery, mutagenesis, amplification, and yeast transformation can be achieved in one to two days. The increased frequency of mutagenesis is applicable to any protein engineering method and is strongly recommended both to improve the speed of binder isolation as well as the overall efficacy. Shuffling can be implemented in any analogous protein scaffold provided the regions of interest are relatively proximal at the DNA level to enable overlap for homologous recombination. Yeast surface display provides an effective system for shuffling because DNA fragments can be recombined with high fidelity during cellular transformation simplifying the method.

The sequence diversity of the highest-affinity clones is striking. L8.5.2 has two cysteines at locations consistent with formation of an inter-loop disulfide bond, as was found previously from a different Fn3 library screened by yeast display[6]. However, unlike in that case, the disulfide is dispensable for high affinity binding, since L7.5.1 has highly related loop sequences but lacks both cysteines. A third clone with similar affinity (L8.5.7, 2.8 pM) was found with a substantially different FG loop, but conserved DE and BC loops.

It is noteworthy that the focused mutagenesis results indicate the potential for mild improvement of affinity maturation through targeted randomization at multiple residues identified as variable through sequence analysis. It is not yet clear if the success of this avenue of affinity maturation resulted from the relatively high number of mutations from the parent clone (eight amino acid changes in the two loops) or the ability to reach amino acids that would require more than one nucleotide mutation. Regardless, the fact that a secondary, semi-rational approach was only able to yield 3.1x enhancement in affinity is indicative of a relatively effective search of sequence space by the initial affinity maturation.

Diversification of the DE loop is an important aspect of Fn3 engineering. As the shortest and least flexible wild-type loop, it is unclear if the binding potential gained by diversification offsets the possible structural destabilization. Thus, analysis of the contribution of the DE loop to binding and stability are valuable to clone maturation and future Fn3 library design. The originally selected DE loop, GVTN, stabilizes clone L7.5.1 relative to wild-type but does not significantly aid binding. Though loop reversion analysis suggests that the binding paratope is dominated by the FG and BC loops, the DE loop can be engineered for improved binding (DE0.4.1) albeit at the expense of stability. Since selections were explicit for affinity, the lack of binding performance from the L7.5.1 DE loop likely resulted from a lack of DE loop diversity after a few rounds of directed evolution (Table 2.2). Thus, a future improvement on loop shuffling will be to incorporate a small percentage of naïve loops with the engineered loops.

A broader uncertainty regarding the DE loop is the extent to which it should be diversified in future Fn3 libraries. Previous binders have been engineered with wild-type DE loops with 350 pM affinity for lysozyme,[6] 30 nM for maltose-binding protein,[5] 250 nM for Src SH3 domain,[2] and approximately 5 μM for ubiquitin.[4] The DE reversion of L7.5.1 represents a 3.7 pM binder with a wild-type DE loop. Collectively, it is clear that binders can be engineered without DE loop diversity. Conversely, engineered DE loops in the L7.5.1 context can either improve the affinity to 1.1 pM or stabilize the molecule by an 11° increase in $T_{1/2}$, the latter of which is not surprising given its extensive contact with the BC loop residues as well as the shortened BC loop length. Moreover, engineered vascular endothelial growth factor receptor 2 binders require their engineered DE loop for binding though it destabilizes the molecule by 2.0 kcal/mol or ~30° decrease in T_m.[7] Consequently, as expected, DE loop impact is context-dependent. One possible general approach would be to introduce mild DE diversity in the original library to allow for binding constraint removal and structural complementation of selected BC loops as well as the possibility of beneficial binding contacts. After binder selection, affinity or stability maturation could be employed with more diverse DE loop shuffling in the context of effective BC and FG loops.

Unlike several examples of previous high affinity Fn3 binder engineering, the selected high affinity clones are relatively stable and monomeric. L7.5.1 is a 2.6 pM binder and

>99% monomeric with a midpoint of thermal denaturation of 56-59°. Moreover, reversion of serine to proline at position 15 improves the $T_{1/2}$ to 60°. In addition, Cons0.4.1 is a 1.1 pM binder with T_m = 53°, and L8.5.7 is a 2.8 pM binder, is 93% monomeric, and has a T_m of 53-55°. Though stability was not an explicit element of selection, the eukaryotic secretion machinery of yeast provides some level of quality control against misfolded proteins.[27; 28] In general, Fn3 clones of higher stability are displayed at high densities on the yeast cell surface (data not shown). Thus, unstable clones are slightly selected against based on the two-color sort regions. It remains to be seen if stable clones will generally result from selections by yeast surface display or if this was a fortuitous result; regardless, yeast surface display provides a means for stability engineering either during or after binder selection.

Overall, the method can be further improved through refinement of naïve loop length distribution, an increase in magnitude of the initial library, inclusion of naïve loops during loop shuffling, and perhaps more expansive mutation than attainable by error-prone PCR with nucleotide analogs. Nevertheless, the combination of yeast surface display, loop length diversity, and dual mode affinity maturation rapidly yielded stable, high affinity binders. The method should be valuable towards development of Fn3 binders to additional targets as well as transferrable to engineering of other proteins for any screenable functionality.

Materials and Methods

Fn3 library construction

Oligonucleotides were purchased from MWG Biotech (High Point, NC) and Integrated DNA Technologies (Coralville, IA). The Fn3 library was constructed to produce wild-type sequence in the framework regions and to randomize the BC, DE, and FG loops of tenth type III domain of human fibronectin. The DNA encoding for amino acids 23-30 (DAPAVTVR) was replaced by $(NNB)_x$ where $x = 6, 7, 8$, or 9 to yield a loop length that is -2, -1, 0, or +1 amino acids relative to wild-type. Similarly, the DNA for amino acids 52-56 (GSKST) was replaced by $(NNB)_y$ where $y = 4, 5, 6$, or 7 and the DNA for amino acids 77-86 (GRGDSPASSK) was replaced by $(NNB)_z$ where $z = 5, 6, 8$, or 10.

The library was constructed by sequential annealing and extension of eight overlapping oligonucleotides.[6] The following components were combined in a 50 μL reaction: two oligonucleotides (0.2 μM a2, $b3n_{mix}$, c6t, or $d7n_{mix}$ + 0.4 μM a1, b4n, $c5n_{mix}$, or d8n, respectively), 1x Polymerase buffer, 0.2 mM dNTPs, 1 mM $MgSO_4$, 1U KOD Hot Start DNA Polymerase (Novagen, Madison, WI), 1M betaine, and 3% DMSO. The mixture was denatured at 95° for 2 min. followed by ten cycles of 94° for 30 s, 58° for 30 s, and 68° for 1 min. and a final extension of 68° for 10 min. Forty μL of the products (a1+a2, $b3n_{mix}$+b4n, $c5n_{mix}$+c6t, or $d7n_{mix}$+d8n) were combined and thermally cycled at identical conditions. Two μL of this product was combined with 0.4 μM primer (a1-b4n amplified with p1; $c5n_{mix}$-d8n amplified with p8) in a new 100 μL reaction and thermally cycled under identical conditions to amplify the appropriate strand. The products were combined and thermally cycled at identical conditions. The final products were concentrated with PelletPaint (Novagen). The plasmid acceptor vector pCTf1f4(Ref. 6) was digested with NcoI, NdeI, and SmaI (New England Biolabs, Ipswich, MA). Multiple aliquots of ~10 μg of Fn3 gene and 3 μg plasmid vector were combined with 50-100 μL of electrocompetent EBY100 and electroporated at 0.54 kV and 25 μF. Homologous recombination of the linearized vector and degenerate insert yielded intact plasmid. Cells were grown in YPD (1% yeast extract, 2% peptone, 2% glucose) for 1 h at 30°, 250 rpm. The number of total transformants was 6.5×10^7 cells as determined by serial dilutions plated on SD-CAA plates (0.1M sodium phosphate, pH 6.0, 182 g/L sorbitol, 6.7 g/L yeast nitrogen base, 5 g/L casamino acids, 20 g/L glucose). The library was propagated by selective growth in SD-CAA, pH 5.3 (0.07M sodium citrate, pH 5.3, 6.7 g/L yeast nitrogen base, 5 g/L casamino acids, 20 g/L glucose, 0.1 g/L kanamycin, 100 kU/L penicillin, and 0.1 g/L streptomycin) at 30°, 250 rpm.

Fluorescence-activated cell sorting

Yeast were grown in SD-CAA, pH 5.3 at 30°, 250 rpm to logarithmic phase, pelleted, and resuspended to 1×10^7 cells/mL in SG-CAA, pH 6.0 (0.1M sodium phosphate, pH 6.0, 6.7 g/L yeast nitrogen base, 5 g/L casamino acids, 19 g/L dextrose, 1 g/L glucose, 0.1 g/L kanamycin, 100 kU/L penicillin, and 0.1 g/L streptomycin) to induce protein expression. Induced cells were grown at 30°, 250 rpm for 12-24h.

Round 0 (three FACS selections) and round 1 (two FACS selections) were conducted with multivalent lysozyme prepared by incubating streptavidin-fluorophore (R-phycoerythrin, AlexaFluor488, or AlexaFluor633, Invitrogen, Carlsbad, CA) with biotinylated lysozyme (Sigma, St. Louis, MO) in a 1:3 ratio in PBSA. Yeast were pelleted, washed in 1 mL PBSA (0.01M sodium phosphate, pH 7.4, 0.137M sodium chloride, 1 g/L bovine serum albumin), resuspended in PBSA with 10-40 mg/L mouse anti-c-myc antibody (clone 9E10, Covance, Denver, PA), and incubated on ice. Cells were washed with 1 mL PBSA and resuspended in PBSA with multivalent lysozyme (0.5 μM for first four sorts and 50 nM for fifth sort) and goat anti-mouse antibody conjugated to R-phycoerythrin, AlexaFluor488, or AlexaFluor633.

Intermediate FACS selections were conducted with near-equilibrium labeling with monovalent lysozyme. Three, two, two, and three selections were performed in rounds 2-5, respectively. Yeast were pelleted, washed in 1 mL PBSA, resuspended in PBSA with biotinylated lysozyme (ranging from 1 μM to 20 pM) and mouse anti-c-myc antibody, and incubated on ice. Cells were then washed with 1 mL PBSA and resuspended in PBSA with streptavidin-fluorophore and fluorophore-conjugated goat anti-mouse antibody.

FACS selections of very high affinity populations were conducted with kinetic competition. Two, three, and two selections were performed in rounds 6-8. Yeast were washed and incubated briefly with 1-2 nM biotinylated lysozyme. Yeast were then washed and resuspended with PBSA with 140 nM unbiotinylated lysozyme (to prevent further association of labeled target) and incubated at room temperature for 2 h to 7 days to enable dissociation of biotinylated lysozyme. Cells were washed in PBSA, resuspended in PBSA with mouse anti-c-myc antibody, and incubated on ice. Cells were washed and labeled with secondary reagents as in equilibrium labeling.

In all cases, labeled cells were washed with 1 mL PBSA, resuspended in 0.5-2.0 mL PBSA and analyzed by flow cytometry using either a MoFlo (Cytomation, Carpinteria, CA) or Aria (Becton Dickinson, Franklin Lakes, NJ) cytometer. C-myc$^+$ cells with the top 0.2-3% of lysozyme binding : c-myc display ratio were selected. Collected cells were grown in SD-CAA, pH 5.3 at 30°, 250 rpm and either induced in SG-CAA, pH 6.0 for further selection or used for plasmid recovery.

Fn3 Mutagenesis

Plasmid DNA from $1x10^8$ cells was isolated using two columns of Zymoprep kit II (Zymo Research, Orange, CA) according to the manufacturer's instructions except for additional centrifugation of neutralized precipitate. The zymoprep elution was cleaned using the Qiagen PCR Purification kit (Qiagen, Valencia, CA), and eluted in 40 µL of elution buffer. Error-prone PCR of the entire Fn3 gene was performed in a 50 µL reaction containing 1x *Taq* buffer, 2 mM $MgCl_2$, 0.5 µM each of primers W5 and W3, 0.2 mM (each) dNTPs, 5 µL of zymoprepped DNA template, 2 mM 8-oxo-dGTP (TriLink, San Diego, CA), 2 mM dPTP (TriLink), and 2.5U of *Taq* DNA polymerase (Invitrogen). In parallel, error-prone PCR of the loop regions was performed via three separate 50 µL reactions with 20 mM 8-oxo-dGTP and 20 mM dPTP and primers BC5 and BC3 for the BC loop, DE5 and DE3 for the DE loop, and FG5 and FG3 for the FG loop. The reactions were denatured at 94° for 3 min., cycled 15 times at 94° for 45 s, 60° for 30 s, and 72° for 90 s, and finally extended at 72° for 10 min. Multiple preliminary mutagenesis reactions of the wild-type plasmid were conducted at different nucleotide analog concentrations. Sequence analysis and comparison to a theoretical framework (Appendix A) indicated the aforementioned conditions yield 1-5 amino acid mutations per gene. The PCR products were purified by agarose gel electrophoresis and each amplified in four 100 µL PCR reactions containing 1x *Taq* buffer, 2 mM $MgCl_2$, 1 µM of each primer, 0.2 mM (each) dNTPs, 4 µL of error-prone PCR product (of 40 µL from gel extraction), and 2.5U of *Taq* DNA polymerase. The reactions were thermally cycled at the same conditions except that 35 cycles were used. Reaction products were concentrated with PelletPaint (Novagen) and resuspended in 1 µL of water.

Plasmid pCT-Fn3 was digested with PstI, BtgI, and BamHI to create linearized vector pCT-Fn3-Gene with the entire Fn3 gene removed. Plasmid pCT-Fn3 was digested with BclI, BtgI, and PasI to create linearized vector pCT-Fn3-Loop with the wild-type gene removed from the BC loop through the FG loop. 100 μL of electrocompetent EBY100 were combined with 0.5-2.0 μg of pCT-Fn3-Gene and the gene-based PCR product and electroporated at 0.54 kV and 25 μF in a 2 mM electroporation cuvette. Similarly, 100 μL of electrocompetent EBY100 were combined with 0.5-2.0 μg of pCT-Fn3-Loop and the three loop-based PCR products and electroporated. Homologous recombination of the linearized vector and mutagenized insert(s) yielded intact plasmid. Cells were grown in YPD for 1 h at 30°, 250 rpm. The medium was switched to SD-CAA to enable selective propagation of successful transformants via growth at 30°, 250 rpm for 24-48 h.

DNA sequencing

Plasmid DNA was isolated using the Zymoprep kit II, cleaned using the Qiagen PCR Purification kit, and transformed into DH5α (Invitrogen) or XL1-Blue *E. coli* (Stratagene, La Jolla, CA). Individual clones were grown, miniprepped, and sequenced using BigDye chemistry on an Applied Biosystems 3730.

Measurement of K_d, k_{on}, and k_{off}

The equilibrium dissociation constant for a clone was determined essentially as described.[25] Briefly, yeast containing the plasmid for a Fn3 clone were grown and induced as for FACS selection. Cells were washed in 1 mL PBSA and resuspended in PBSA containing biotinylated lysozyme in concentrations generally spanning four orders of magnitude surrounding the equilibrium dissociation constant. The number of cells and sample volumes were selected to ensure excess lysozyme relative to Fn3. For clones of low picomolar affinity, this criterion necessitates very low cell density, which makes cell collection by centrifugation procedurally difficult. To obviate this difficulty, uninduced cells are added to the sample to enable effective cell pelleting during centrifugation with no effect on lysozyme binding of the Fn3-displaying induced cells. Cells were incubated at 25° for sufficient time to ensure that the approach to equilibrium was at least 98% complete. Cells were then pelleted, washed with 1 mL PBSA, and incubated in PBSA with 10 mg/L streptavidin-R-phycoerythrin for 10-30 min. Cells were washed and resuspended with PBSA and analyzed with an Epics XL flow cytometer. The minimum and maximum fluorescence and the K_d value were determined by minimizing the sum of squared errors.

For determination of the dissociation constant, k_{off}, clonal cell cultures were grown, induced, and washed as above. Cells were incubated in PBSA with a saturating concentration of biotinylated lysozyme at 25°. At various times, an aliquot of cells was washed with PBSA with excess unbiotinylated lysozyme, resuspended in PBSA with excess unbiotinylated lysozyme, and incubated at 25°. Simultaneously, all samples of differing dissociation times were washed with PBSA and incubated in 10 mg/L streptavidin-R-phycoerythrin for 10-30 min. Cells were washed and resuspended with PBSA and analyzed with an Epics XL flow cytometer. The minimum and maximum fluorescence and the k_{off} value were determined by minimizing the sum of squared errors.

For determination of the association constants, k_{on}, clonal cell cultures were grown, induced, and washed as above. At various times, an aliquot of cells was resuspended in biotinylated lysozyme and incubated at 25°. Simultaneously, all samples of differing association times were washed with PBSA with excess unbiotinylated lysozyme and incubated in PBSA with 10 mg/L streptavidin-R-phycoerythrin for 10-30 min. Cells were washed and resuspended with PBSA and analyzed with an Epics XL flow cytometer. The maximum fluorescence and k_{on} were determined by minimizing the sum of squared errors assuming a 1:1 binding model. The experimentally determined value of k_{off} was used to determine the effective association rate, $k_{on}[Lysozyme] + k_{off}$.

The equilibrium dissociation constant was also determined for the soluble forms of L7.5.1 and Cons0.4.1 by equilibrium competition titration. Varying concentration of purified Fn3 domains were incubated with 20 pM biotinylated lysozyme in 50 mL of PBSA. Yeast displaying L7.5.1 were added and incubated for 7 days to near equilibrium. Cells were then pelleted, washed with 1 mL PBSA, and incubated in PBSA with 10 mg/L streptavidin-R-phycoerythrin for 15 min. Cells were washed and resuspended with PBSA and analyzed with an Epics XL flow cytometer. A two-state binding model was assumed and the minimum and maximum fluorescence and equilibrium dissociation constant were determined by minimizing the sum of squared errors.

Fn3 Production and Biophysical Characterization

Fn3 clones were produced as previously described.[6] Briefly, BL21(DE3)pLysS *E. coli* (Invitrogen) containing the pET-24b-based Fn3 plasmid were grown in Luria-Bertani medium with 50 μg/mL kanamycin and 34 μg/mL chloramphenicol at 37°, 250 rpm to an A_{600} of 0.1-0.2 and induced with 0.5 mM IPTG for 18-24 h at 30°, 250 rpm. Cells were lysed by sonication and the insoluble fraction was removed by centrifugation at 19,000*g* for 40 min. His_6-tagged Fn3 was purified from the soluble fraction with TALON Superflow Metal Affinity Resin (Clontech, Mountain View, CA), dialyzed against PBS, and concentrated to 0.5 mL with an Amicon Ultra centrifugal filter (Millipore, Billerica, MA).

The oligomeric state was analyzed by size exclusion chromatography on a Superdex 75 HR10/300 column (Amersham Pharmacia Biotech, Piscataway, NJ). Monomer was isolated for biophysical analysis. PBS standards or monomeric protein in PBS was thermally denatured from 25° to 95° at a rate of 1°/min. in a differential scanning calorimeter (VP-DSC, MicroCal). Irreversible aggregation of L7.5.1 occurs at high temperatures. The midpoint of thermal denaturation for this clone is identified as the temperature of maximum heat capacity. Samples were dialyzed in 10 mM sodium phosphate buffer, pH 7.0 and diluted to 8-10 μM for far-UV circular dichroism analysis. Ellipticity was measured from 250-190 nm on an Aviv 202 spectrometer (Aviv Biomedical, Lakewood, NJ) with a quartz cuvette with a 1 mm path length (New Era, Vineland, NJ). Thermal denaturation was conducted by measuring ellipticity at 216 nm from 25° to 95° and calculating T_m from a standard two-state unfolding curve.

Thermal stabilities were also determined using a yeast surface display thermal denaturation assay derived from Orr, *et al.*[30] Fn3 was displayed on the yeast surface as for measurement of kinetic and equilibrium binding constants. Cells were washed and resuspended with PBSA, incubated at 20-85° for 30 min., and incubated on ice for 5 min.

Biotinylated lysozyme was added at a saturating concentration (*e.g.*, 20 nM for L7.5.1) and mouse anti-c-myc antibody was added at 40 mg/L and incubated on ice for 20 min. Cells were washed and incubated in PBSA with 10 mg/L streptavidin-R-phycoerythrin and 25 mg/L AlexaFluor488 conjugated goat anti-mouse antibody. Cells were washed and resuspended in PBSA and analyzed on an Epics XL flow cytometer. The minimum and maximum fluorescence (F_{min} and F_{max}), the $T_{1/2}$, and the enthalpy of unfolding at $T_{1/2}$ (ΔH_m) were determined by minimizing the sum of squared errors between experimental data and theoretical values according to a two-state unfolding equation:

$$\frac{F - F_{\min}}{F_{\max} - F_{\min}} = f_{folded} = \left\{1 + \exp\left[\frac{\Delta H_m}{R}\left(\frac{1}{T_m} - \frac{1}{T}\right)\right]\right\}^{-1}$$

L7.5.1 Reversion Clone Construction

Reversion of engineered loops of L7.5.1 to wild-type sequence was accomplished by annealing and extending two PCR products created with a gene-terminal primer and a primer that annealed adjacent to the loop of interest but was extended to include wild-type sequence. Specifically, one PCR reaction contained a 5’ gene terminal primer and a primer that annealed to the 25 nucleotides immediately 5’ of the loop of interest but included a non-annealing ‘tail’ encoding for the wild-type loop sequence. In parallel, PCR was performed with a 3’ gene terminal primer and a primer annealing to the 25 nucleotides immediately 3’ of the loop and including a non-annealing tail. The first PCR product encodes from the start of the gene to the loop of interest and the second PCR product encodes from the loop of interest to the end of the gene. These two products are annealed and extended to yield the full Fn3 gene containing the wild-type sequence in the loop of interest. Framework reversions were introduced by standard site-directed mutagenesis using the QuikChange Mutagenesis Kit (Stratagene) according to the manufacturer's instructions. Clone construction was verified by DNA sequencing.

Focused library construction

The DE randomization library was created in a similar manner to the L7.5.1 loop reversion clones. One PCR amplified the L7.5.1 S15P gene fragment 5' of the DNA encoding for the DE loop. A second PCR amplified the gene 3' of the DNA encoding the

DE loop using a primer that included a degenerate (NNB) DE loop sequence and 20 nucleotides of overlap with the other PCR product. The PCR products were annealed, extended to produce the full gene, and amplified. This process was conducted independently with four oligonucleotides encoding the four different DE loop lengths. The gene fragments were electroporated into electrocompetent EBY100 along with pCT-Fn3-Loop vector. The resulting library encoded for L7.5.1 S15P with a fully random DE loop of length 4, 5, 6, or 7 amino acids.

A library randomizing the unconserved residues of clones similar to L7.5.1 was constructed by PCR of L7.5.1 S15P. The 5' primer contained 17 nucleotides 5' of the BC loop, 21 nucleotides to encode the BC loop, and 22 nucleotides to anneal 3' of the BC loop. The 3' primer contained 19 nucleotides 3' of the FG loop, 30 nucleotides to encode for the FG loop, and 10 nucleotides 5' of the FG loop (note that the nucleotides encoding the first three conserved amino acids of the FG loop also enable annealing during PCR). The PCR products were amplified with extended primers to increase the length of the conserved sequence flanking the loops to improve homologous recombination. The gene fragments were electroporated into electrocompetent EBY100 along with pCT-Fn3-Loop vector. Two versions of the BC and FG loops were included. One oligonucleotide completely randomized the unconserved residues using NNB degeneracy (BC: RXXPWAX; FG: RVGRXXXXXG). The other oligonucleotide restricted diversity to amino acids observed during affinity maturation (BC: R(D/G)(C/H/R/Y)PWA(I/T); FG: RVG(R/W)(A/M/T/V)(F/L/P/S)(C/D/G/Y)(A/T)(L/P/S)(G/S).

References

1. Huang, J., Koide, A., Nettle, K. W., Greene, G. L. & Koide, S. (2006). Conformation-specific affinity purification of proteins using engineered binding proteins: Application to the estrogen receptor. *Protein Expression and Purification* **47**, 348-354.
2. Karatan, E., Merguerian, M., Han, Z. H., Scholle, M. D., Koide, S. & Kay, B. K. (2004). Molecular recognition properties of FN3 monobodies that bind the Src SH3 domain. *Chemistry & Biology* **11**, 835-844.
3. Koide, A., Abbatiello, S., Rothgery, L. & Koide, S. (2002). Probing protein conformational changes in living cells by using designer binding proteins: Application to the estrogen receptor. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 1253-1258.
4. Koide, A., Bailey, C. W., Huang, X. L. & Koide, S. (1998). The fibronectin type III domain as a scaffold for novel binding proteins. *Journal of Molecular Biology* **284**, 1141-1151.

5. Koide, A., Gilbreth, R. N., Esaki, K., Tereshko, V. & Koide, S. (2007). High-affinity single-domain binding proteins with a binary-code interface. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 6632-6637.
6. Lipovsek, D., Lippow, S. M., Hackel, B. J., Gregson, M. W., Cheng, P., Kapila, A. & Wittrup, K. D. (2007). Evolution of an interloop disulfide bond in high-affinity antibody mimics based on fibronectin type III domain and selected by yeast surface display: Molecular convergence with single-domain camelid and shark antibodies. *Journal of Molecular Biology* **368**, 1024-1041.
7. Parker, M. H., Chen, Y., Danehy, F., Dufu, K., Ekstrom, J., Getmanova, E., Gokemeijer, J., Xu, L. & Lipovsek, D. (2005). Antibody mimics based on human fibronectin type three domain engineered for thermostability and high-affinity binding to vascular endothelial growth factor receptor two. *Protein Engineering Design & Selection* **18**, 435-444.
8. Richards, J., Miller, M., Abend, J., Koide, A., Koide, S. & Dewhurst, S. (2003). Engineered fibronectin type III domain with a RGDWXE sequence binds with enhanced affinity and specificity to human alpha v beta 3 integrin. *Journal of Molecular Biology* **326**, 1475-1488.
9. Xu, L. H., Aha, P., Gu, K., Kuimelis, R. G., Kurz, M., Lam, T., Lim, A. C., Liu, H. X., Lohse, P. A., Sun, L., Weng, S., Wagner, R. W. & Lipovsek, D. (2002). Directed evolution of high-affinity antibody mimics using mRNA display. *Chemistry & Biology* **9**, 933-942.
10. Rock, E., Sibbald, P., Davis, M. & Chien, Y. (1994). CDR3 length in antigen-specific immune receptors. *J. Exp. Med.* **179**, 323-328.
11. Ohlin, M. & Borrebaeck, C. A. K. (1996). Characteristics of human antibody repertoires following active immune responses in vivo. *Molecular Immunology* **33**, 583-592.
12. Lamminmaki, U., Pauperio, S., Westerlund-Karlsson, A., Karvinen, J., Virtanen, P. L., Lovgren, T. & Saviranta, P. (1999). Expanding the conformational diversity by random insertions to CDRH2 results in improved anti-estradiol antibodies. *Journal of Molecular Biology* **291**, 589-602.
13. Lee, C. V., Liang, W. C., Dennis, M. S., Eigenbrot, C., Sidhu, S. S. & Fuh, G. (2004). High-affinity human antibodies from phage-displayed synthetic fab libraries with a single framework scaffold. *Journal of Molecular Biology* **340**, 1073-1093.
14. Batori, V., Koide, A. & Koide, S. (2002). Exploring the potential of the monobody scaffold: effects of loop elongation on the stability of a fibronectin type III domain. *Protein Engineering* **15**, 1015-1020.
15. VanAntwerp, J. J. & Wittrup, K. D. (2000). Fine affinity discrimination by yeast surface display and flow cytometry. *Biotechnology Progress* **16**, 31-37.
16. Bowley, D. R., Labrijn, A. F., Zwick, M. B. & Burton, D. R. (2007). Antigen selection from an HIV-1 immune antibody library displayed on yeast yields many novel antibodies compared to selection from the same library displayed on phage. *Protein Engineering Design & Selection* **20**, 81-90.
17. Ling, M. M. (2003). Large antibody display libraries for isolation of high-affinity antibodies. *Combinatorial Chemistry & High Throughput Screening* **6**, 421-432.
18. Sheets, M. D., Amersdorfer, P., Finnern, R., Sargent, P., Lindqvist, E., Schier, R., Hemingsen, G., Wong, C., Gerhart, J. C. & Marks, J. D. (1998). Efficient construction of a large nonimmune phage antibody library: The production of high-affinity human single-chain antibodies to protein antigens. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 6157-6162.
19. Hanes, J., Schaffitzel, C., Knappik, A. & Pluckthun, A. (2000). Picomolar affinity antibodies from a fully synthetic naive library selected and evolved by ribosome display. *Nature Biotechnology* **18**, 1287-1292.
20. Stemmer, W. P. C. (1994). Rapid Evolution of a Protein in-Vitro by DNA Shuffling. *Nature* **370**, 389-391.
21. Jirholt, P., Ohlin, M., Borrebaeck, C. A. K. & Soderlind, E. (1998). Exploiting sequence space: shuffling in vivo formed complementarity determining regions into a master framework. *Gene* **215**, 471-476.
22. Marks, J. D., Griffiths, A. D., Malmqvist, M., Clackson, T. P., Bye, J. M. & Winter, G. (1992). Bypassing Immunization - Building High-Affinity Human-Antibodies by Chain Shuffling. *Bio-Technology* **10**, 779-783.
23. Barbas, C. F., Hu, D., Dunlop, N., Sawyer, L., Cababa, D., Hendry, R. M., Nara, P. L. & Burton, D. R. (1994). In-Vitro Evolution of a Neutralizing Human-Antibody to Human-

Immunodeficiency-Virus Type-1 to Enhance Affinity and Broaden Strain Cross-Reactivity. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 3809-3813.

24. Cadwell, R. C. & Joyce, G. F. (1992). Randomization of genes by PCR mutagenesis. *PCR Methods Appl* **2**, 28-33.
25. Chao, G., Lau, W. L., Hackel, B. J., Sazinsky, S. L., Lippow, S. M. & Wittrup, K. D. (2006). Isolating and engineering human antibodies using yeast surface display. *Nat. Protocols* **1**, 755-768.
26. Zaccolo, M., Williams, D. M., Brown, D. M. & Gherardi, E. (1996). An approach to random mutagenesis of DNA using mixtures of triphosphate derivatives of nucleoside analogues. *Journal of Molecular Biology* **255**, 589-603.
27. Kowalski, J. M., Parekh, R. N., Mao, J. & Wittrup, K. D. (1998). Protein folding stability can determine the efficiency of escape from endoplasmic reticulum quality control. *Journal of Biological Chemistry* **273**, 19453-19458.
28. Shusta, E. V., Kieke, M. C., Parke, E., Kranz, D. M. & Wittrup, K. D. (1999). Yeast polypeptide fusion surface display levels predict thermal stability and soluble secretion efficiency. *Journal of Molecular Biology* **292**, 949-956.
29. Moore, G. L. & Maranas, C. D. (2000). Modeling DNA mutation and recombination for directed evolution experiments. *Journal of Theoretical Biology* **205**, 483-503.
30. Orr, B. A., Carr, L. M., Wittrup, K. D., Roy, E. J. & Kranz, D. M. (2003). Rapid method for measuring ScFv thermal stability by yeast surface display. *Biotechnology Progress* **19**, 631-638.
31. Main, A. L., Harvey, T. S., Baron, M., Boyd, J. & Campbell, I. D. (1992). The three-dimensional structure of the tenth type III module of fibronectin: An insight into RGD-mediated interactions. *Cell* **71**, 671-678.

3. CONSTRAINED PROTEIN DIVERSITY WITH A TYROSINE/SERINE CODE AND DE LOOP BIAS

Introduction

Ultra high affinity binding is enabled with three-loop diversification of the Fn3 scaffold, but the resultant expansion of theoretical sequence space necessitates efficient library design. Early molecular recognition library designs incorporated all twenty natural amino acids. However, it has been demonstrated that tyrosine can dominate molecular recognition in antibody domains[1; 2] and that a serine/tyrosine binary code is sufficient to generate nanomolar affinity interactions in an antigen binding fragment.[3] This library approach has been successfully applied to generate low- to mid-nanomolar binders with the Fn3 scaffold.[4] Yet while the binary amino acid code greatly reduces theoretical sequence space enabling isolation of functional clones, it also reduces the potential structural and chemical complementarity of the binder-target interaction. As such, it is not clear which library design is superior for the generation of high affinity binders. A hybrid design with all twenty amino acids yet bias towards tyrosine, serine, and glycine was investigated in an antibody library;[5] this library yielded more high affinity binders than a strictly serine/tyrosine library. This comparison was performed with an antibody domain binding to a single target. The current study directly compares full diversity to serine/tyrosine diversity in the context of the Fn3 domain in two binder engineering campaigns.

In parallel with the study of library design, we sought to develop useful reagents with advantageous biophysical properties. High-yield bacterial expression enables inexpensive production of Fn3 domains. The absence of lysines near the engineered binding surface and the cysteine-free sequence permit both amine- and thiol-based conjugation chemistries for immobilization or fluorophore coupling. The small, single-domain architecture facilitates multifunctional protein fusions. Engineered Fn3 domains have demonstrated utility as detection agents. Biotinylated $\alpha_v\beta_3$ integrin binders were used as a primary label in flow cytometry.[6] Alkaline phosphatase fusions of Src-binding Fn3 domains were effective in Western blotting.[7] The Src binders were also effective in pull-down experiments, and estrogen receptor binders were effective in affinity chromatography.[8] Thus, Fn3 domains that bind immunoglobulin G (IgG) may be broadly useful in these applications as well as in protein microarrays and as adaptors for nanoparticles or other supramolecular assemblies. In this study, high affinity binders to both goat and rabbit IgG were engineered. The resultant binders are characterized in terms of affinity, stability, and specificity and are demonstrated as useful reagents for purification and detection. In addition, both library designs were employed in the generation of binders to carcinoembryonic antigen (Appendix B).

Results

Library Design and Construction

Two libraries, NNB and YS, were constructed for comparison. In both libraries the BC, DE, and FG loop sequences of Fn3 were diversified. The NNB library has been previously described.[9] Eight, five, and ten residues in the BC, DE, and FG loops were randomized using NNB codons; four loop lengths, selected based on phylogenetic occurrence, were included in each loop. The YS library diversified nine, five, and ten residues in the BC, DE, FG loops. The BC and FG loops were randomized between serine and tyrosine. BC loop diversity was extended to include Y31 because it is a large side chain with potential steric conflicts and inclusion of serine only increases the theoretical diversity twofold. The DE loop was diversified with a wild-type bias to improve structural integrity while enabling some diversity to either eliminate detrimental interactions or provide beneficial interactions. Biased nucleotides were synthesized to yield a library with approximately 50% wild-type and 50% of the other 19 amino acids at G52, S53, S55, and T56. K54 was randomized using an NNB codon because the large, charged side chain is potentially sterically and electrostatically detrimental. Four loop lengths were allowed in each loop. Library design is summarized in Table 3.1.

The libraries of Fn3 genes were transformed into a yeast surface display library by homologous recombination with a vector including an N-terminal Aga2p protein for yeast surface tethering to Aga1p and a C-terminal c-myc epitope for full-length Fn3 detection. Electroporation yielded 21.5×10^7 and 25×10^7 transformants from the NNB and YS libraries, respectively. Sequencing and flow cytometry analysis (data not shown) indicate 34% and 60% of clones encode for full-length protein resulting in 7.3×10^7 and 15×10^7 Fn3 in each library.

Table 3.1. *Library Designs*

Library	BC Loop	DE Loop	FG Loop
WT	DAPAVTVRY	GSKST	GRGDSPASSK
NNB	$X_{6\text{-}9}Y$	$X_{4\text{-}7}$	$X_{5,6,8,10}$
YS	$(S\|Y)_{7\text{-}10}$	$gsX_{0,1,3}st$	$(S\|Y)_{6,7,8,10}$

WT: wild-type sequence; *X*: any amino acid; *S|Y*: either serine or tyrosine; *g*: 50% glycine, 50% any other amino acid (analogous for *s, t*); subscripts refer to number of amino acids

Binder Engineering

The efficacies of the NNB and YS libraries were compared for their ability to generate binders to goat IgG and rabbit IgG. The libraries were pooled to enable direct clone competition and eliminate any experimental bias. The libraries were sorted twice for binding to biotinylated IgG immobilized on streptavidin-coated magnetic beads followed by a fluorescence-activated cell sort (FACS) for c-myc$^+$ clones that represent full-length Fn3. The selected population was diversified by both full gene mutagenesis and focused loop mutagenesis with shuffling as described[9]. The transformed mutants, mixed with the original clones, underwent two bead selections and a FACS followed by mutagenesis. Two bead selections were followed by a FACS for both c-myc and binding to biotinylated IgG, detected by streptavidin-fluorophore. Two additional rounds of mutagenesis and FACS with decreasing IgG concentrations were performed. At this point, the population sorted for binding to goat IgG exhibited binding to 500 pM goat IgG; the population sorted for binding to rabbit IgG exhibited binding to 50 pM rabbit IgG (data not shown).

Each intermediate population used for mutagenesis was also sorted three or four additional times without mutagenesis to identify the best clones during each round of affinity maturation. The naïve library was excluded from this analysis because the populations did not exhibit substantial binding until after mutagenesis. Several clones from each final population and two intermediate populations were sequenced (Table 3.2). Both the goat and rabbit IgG binder engineering yielded one dominant clone after four rounds of mutagenesis although extensive sequence diversity appears in intermediate populations.

Table 3.2. *Binder Sequences*

Goat IgG Binders

Clone	#	BC Loop	DE Loop	FG Loop	Framework
WT	-	DAPAVTVR	GSKST	GRGDSPASSK	-
Round 1					
gI1.5.1	1	ALPRSE	GIRS	AHKSVL	S2P, T58I
gI1.5.2	1	HSYYSYY	GFYST	MDGASPLQ	D7G
gI1.5.3	2	KMRAAR	RFRS	GDGHGG	T58I
gI1.5.4	1	NLEIFPR	GIRS	RTRVI	P51S, T58I
gI1.5.5	1	NLGIFPR	GIRS	RTRVI	T58I
Round 2					
gI2.5.1	1	TNLSSS	NWTS	SYGLVISN	T58I, V68A
gI2.5.2	1	ALPRSE	NWTS	SPGLVLGA	T58I
gI2.5.3	2	TRAYFAP	GSLSS	SYGLVITD	P51S, T58I, I88T
gI2.5.4	2	YSSYSYY	GFRPT	YYSSSYY	T35A, E38G, S89P
gI2.5.5	1	RMPVTD	truncation		-
gI2.5.8	1	RLPRSA	NWTS	SPGLILGA	T58I, I90T
gI2.5.9	1	YCSYSYY	GFRSG	FDGVAF	-
Round 3					
gI3.2.1	2	TARMRSP	NWTS	SPGLILGA	T58I
gI3.2.2	2	ALPRSE	NWTS	SPGLVLGA	T58I
gI3.5.1	5	TRAYFAP	GSLSS	SYGLVITD	P51S, T58I, I88T
Round 4					
gI4.5.1	5	TRAYFAP	GSLSS	SYGLVITD	P51S, T58I, I88T

Rabbit IgG Binders

Clone	#	BC Loop	DE Loop	FG Loop	Framework
WT	-	DAPAVTVR	GSKST	GRGDSPASSK	-
Round 1					
rI1.5.1	1	VRPSYSRL	STATT	GYGGRRVQ	P51A, G61R, Y73C
rI1.5.2	1	ATTGKAPL	KGATA	SYYDYHS	Q46K, S60R
rI1.5.3	4	VATSCL	ATWVK	HYDDTLS	-
rI1.5.4	2	ATTGKTPL	RSAEM	HYDDTLS	-
rI1.5.8	1	VATSCL	ATWVK	HYDDTLS	N42G
rI1.5.9	2	ARASNPCL	ATWVK	GYGGKRVQ	Q46R, S60R, K98E
Round 2					
rI2.5.1	1	VRSPYRRL	RSARS	GYGGRRVQ	-
rI2.5.2	2	VNGDSCL	ATWVK	GYGGKRVQ	S60R
rI2.5.3	1	VRPSYSRL	PTHFF	GYGGKRVQ	S60R
rI2.5.4	1	ARPSYSRL	ATWVK	GYGGKRVQ	-
rI2.5.5	2	VRPSYSRL	ATWVK	GYGGKRVQ	-
rI2.5.6	1	VRPSYSRL	KGATV	GYGGKRVQ	P51S
rI2.5.9	1	VRPSYSRL	ATWVK	GYGGERVQ	-
rI2.5.10	1	VATSCL	RSATS	GYGGKRVQ	P51S
Round 3					
rI3.6.2	1	VRPSYSRL	RSWTS	GYGGKRVP	S60R
rI3.6.3	2	SRARNACL	ATWVK	GYGGKRVQ	G61R
rI3.6.4	2	VRPSYSRL	RSARS	GYGGERVQ	P51S, S60R
rI3.6.5	1	AHAPNPCL	ATWVK	GYGGKRVQ	S60R, I90T
rI3.6.6	2	ATTGKAPL	KGATA	SYYDYHS	Q46K, S60R
rI3.6.10	1	ARPSYSRL	GSAHV	GYGGKRVQ	P51S, S60R, G61R
Round 4					
rI4.3.1	2	ATTGKTPL	RSAEM	HYDDTLS	-
rI4.3.3	1	ARASNPCL	ATWVK	GYGGKRVQ	S60R
rI4.3.4	1	VNGDSCL	GSAHV	GYGGKRVQ	P51S, S60R, G61R, A74T
rI4.3.5	1	AHAPNPCL	AAWVE	GYGGKRVQ	S60R
rI4.5.1	4	ATTGKAPL	ATWVK	HYDDTLS	S60R, K98E
rI4.5.5	1	ATTGKAPL	ATWVK	HYDDTLS	S60R

Clones are named as *(g/r)Ix.y.z* where *g* or *r* indicate goat or rabbit species specificity, *x* is the number of mutagenesis steps, *y* is the number of selections after the most recent mutagenesis and *z* is the clone number in that population. # indicates the frequency of occurrence of the indicated clone.

Sequence analysis reveals a strong preference for the NNB library. Since loop shuffling during mutagenesis can recombine loops from multiple sources, loop sequences were analyzed individually. Thirty-nine of 42 clones (93%) have BC loops of NNB origin. Ninety-three percent have FG loops of NNB origin. The DE loop, which was fully randomized in the NNB library and biased towards wild-type in the YS library, exhibited less preference; 65% of clones had DE loops more likely to originate from NNB whereas 23% were more likely to originate from the wild-type bias of YS and 13% were ambivalent. In fact, only a single clone, gI2.5.4, retained all three loops of apparent YS library origin.

Framework Mutations

The dominant clones, gI2.5.3 (which is identical to gI3.5.1 and gI4.5.1) and rI4.5.1, each contain multiple framework mutations (Figure 3.1A,B). The impact of these mutations was investigated in terms of binding and stability. Each framework mutation in the two clones was individually reverted to the wild-type side chain and clonal cultures were assayed for stability and binding. The relative number of Fn3 molecules displayed on the yeast surface after induction at 37°, which correlates to protein stability,[10] was determined by flow cytometry. Binding to 10 nM goat IgG and 100 pM rabbit IgG was assayed by flow cytometry. None of the gI2.5.3 framework mutations significantly impact stability (Figure 3.1) but two of three are important for binding. Reversion of isoleucine to threonine at amino acid 58 ablates binding at 10 nM IgG while reversion of serine to proline at position 51 decreases binding five-fold. Conversely, threonine and isoleucine at position 88 yield indistinguishable results.

The E98K reversion of rI4.5.1, which was also observed as clone rI4.5.5, does not significantly affect binding or stability. The R60S reversion maintains stability but has slightly decreased binding.

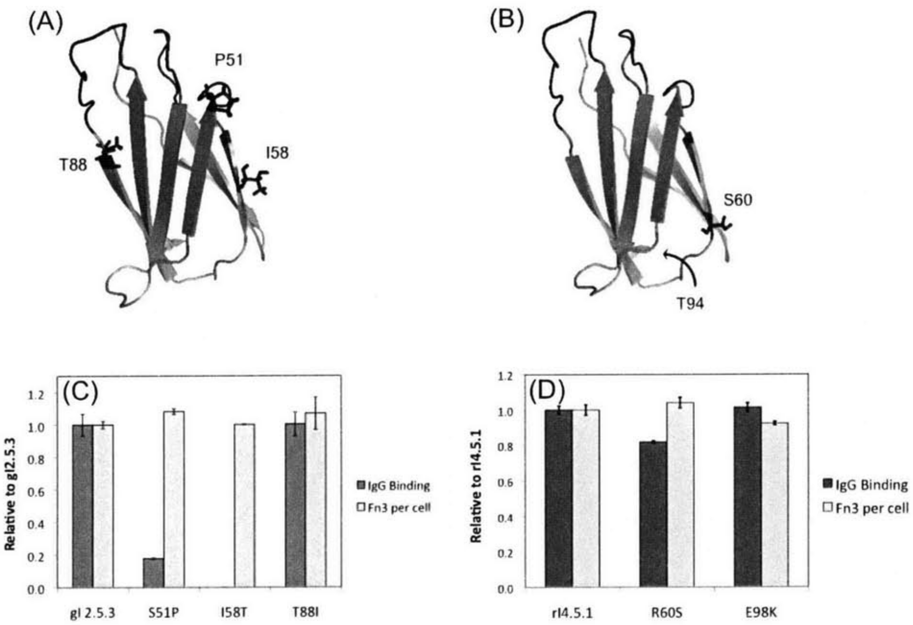


Figure 3.1. *gI2.5.3 and rI4.5.1 framework mutations.* (A) P51, T58, and I88 are presented in the wild-type Fn3 structure 1TTG[11]. (B) S60 is presented in the wild-type Fn3 structure. E98 was not included in this structure, but the C-terminal amino acid of the structure, T94, is indicated. (C) The indicated clone, with a c-myc epitope tag, was displayed on the surface of yeast with 37° induction. Cells were labeled with chicken anti-c-myc antibody followed by bovine anti-chicken phycoerythrin conjugate and analyzed by flow cytometry. The phycoerythrin signals above background were normalized to the gI2.5.3 sample and are presented as *Fn3 per cell.* Cells were labeled with 10 nM goat IgG-FITC conjugate and analyzed by flow cytometry. The FITC signal above background was divided by the relative number of Fn3 per cell, normalized to the gI2.5.3 sample, and is presented as *IgG Binding.* Values are the mean ± standard deviation of at least duplicate experiments. (D) As in (C) except samples were labeled with 100 pM rabbit IgG and were normalized to rI4.5.1.

Affinity Analysis

The affinities of several clones were determined by titration using yeast surface display and soluble IgG. Although avidity effects resulting from multivalent display of Fn3 binding to homodimeric IgG can enable improved binding at low concentrations relative to a monovalent interaction, it has been demonstrated that the concentration of half-maximal binding corresponds to the monovalent equilibrium dissociation constant, K_d. The highest affinity goat IgG binder, gI2.5.3T88I, has a K_d of 1.2 ± 0.4 nM (Figure 3.2, Table 3.3). Clones gI2.5.2 and gI2.5.4, the YS clone, have mid-nanomolar affinities. The highest affinity rabbit IgG binder, rI4.5.5, exhibits 51±4 pM affinity (Figure 3.2, Table 3.3). Other clones demonstrate picomolar to low nanomolar affinities.

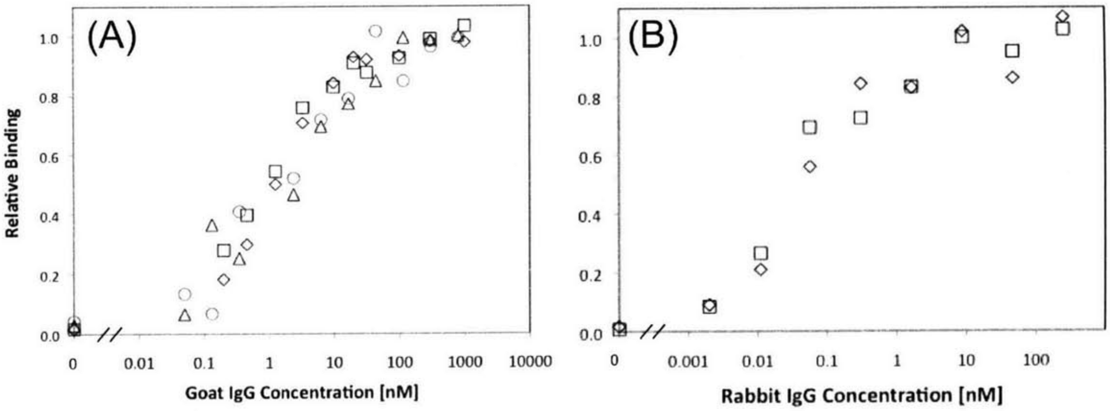


Figure 3.2. *Affinity titrations of gI2.5.3T88I and rI4.5.5.* Yeast displaying gI2.5.3T88I (A) or rI4.5.5 (B) were incubated at 22° with the indicated concentration of FITC-conjugated goat IgG or AlexaFluor633-conjugated rabbit IgG. The cells were washed and analyzed by flow cytometry. The signals relative to non-displaying yeast were normalized to the maximum signal.

Table 3.3. *Binding Affinity*

Clone	Amino Acid Sequence				K_d [nM]
	BC	DE	FG	Framework	
			Goat IgG Binders		
gI2.5.3T88I	TRAYFAP	GSLSS	SYGLVITD	P51S, T58I	1.2 ± 0.4
gI2.5.2	ALPRSE	NWTS	SPGLVLGA	T58I	32 ± 21
gI2.5.4	YSSYSYY	GFRPT	YYSSSYY	T35A, E38G, S89P	35 ± 16
			Rabbit IgG Binders		
rI4.5.5	ATTGKAPL	ATWVK	HYDDTLS	S60R	0.051 ± 0.004
rI4.3.1	ATTGKTPL	RSAEM	HYDDTLS	-	0.117 ± 0.006
rI3.6.6	ATTGKAPL	KGATA	SYYDYHS	Q46K, S60R	0.187 ± 0.034
rI4.3.4	VNGDSCL	GSAHV	GYGGKRVQ	P51S, S60R, G61R, A74T	0.30 ± 0.12
rI3.6.4	VRPSYSRL	RSARS	GYGGERVQ	P51S, S60R	0.63 ± 0.07
rI4.3.3	ARASNPCL	ATWVK	GYGGKRVQ	S60R	1.08 ± 0.38

Stability Analysis

Thermal stabilities were analyzed using a previously validated yeast surface display assay.[9; 12] The highest affinity binders, gI2.5.3T88I and rI4.5.5, have midpoints of thermal denaturation of 63.9 ± 0.3° and 49.1 ± 0.5°, respectively (Figure 3.3). The serine/tyrosine clone gI2.5.4 has a $T_{1/2}$ of 57.3 ± 0.9°. rI4.3.4, the second highest affinity rabbit IgG binder without sequence homology to rI4.5.5, has a $T_{1/2}$ of 44.8 ± 1.9°. gI2.5.2 has a midpoint of denaturation of 59.3 ± 2.5°. rI3.6.4, which has FG loop homology to rI4.3.4 but novel BC and DE loops, has a $T_{1/2}$ of 49.5 ± 0.5°.

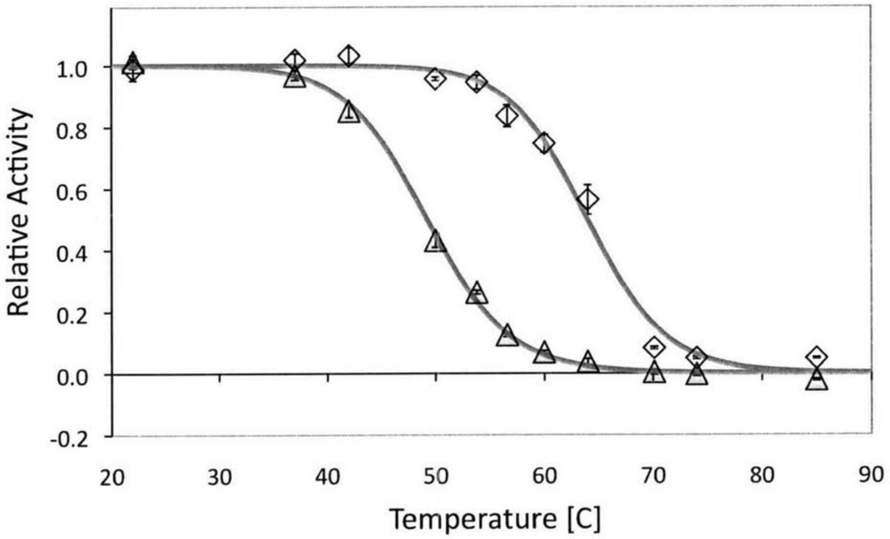


Figure 3.3. *Thermal stability.* Yeast displaying gI2.5.3T88I (diamonds) or rI4.5.5 (triangles) were incubated at the indicated temperature for 30 minutes, returned to ice, labeled with FITC-conjugated IgG, and analyzed by flow cytometry. The FITC signal relative to non-displaying cells was normalized by the 22° sample. Data represent the mean ± standard deviation of triplicate samples.

Specificity Analysis

The specificity of binding was assayed by flow cytometry. None of the tested clones bind the non-cognate proteins lysozyme or streptavidin (Table 3.4). rI3.6.4, rI4.3.4, and rI4.5.5 exhibit unique specificity for rabbit IgG with no detectable binding to bovine, chicken, goat, human, or mouse IgG. Conversely, all three goat IgG binders tested also bind bovine IgG. Clone gI2.5.3T88I also binds mouse IgG.

Table 3.4. *Binding specificity.*

Target	gI2.5.2	gI2.5.3T88I	gI2.5.4	rI3.6.4	rI4.3.4	rI4.5.5
Bovine IgG	+	+	+	-	-	-
Chicken IgG	-	-	-	-	-	-
Goat IgG	+	+	+	-	-	-
Human IgG	-	-	-	-	-	-
Mouse IgG	-	+	-	-	-	-
Rabbit IgG	-	-	-	+	+	+
Lysozyme	-	-	-	-	-	-
Streptavidin	-	-	-	-	-	-

Yeast displaying the indicated clone were labeled with 100 nM protein and analyzed by flow cytometry. The presence (+) or absence (-) of binding is indicated.

Utility in Affinity Purification

The utility of the engineered Fn3 domains in affinity purification from complex mixtures was investigated. Fn3 domains were produced in *E. coli* and biotinylated on exposed amines. To increase biotinylation potential, a two-lysine tail was included. Also, the two engineered lysines in rI4.5.5 were reverted to serine (rI4.5.5K27S/K56S) to avoid biotinylation of the paratope. Affinity analysis reveals that these mutations only increase the K_d threefold thereby maintaining picomolar affinity.

Goat or rabbit serum was applied to a column containing biotinylated Fn3 (gI2.5.3T88I or rI4.5.5K27S/K56S) immobilized on streptavidin agarose. The column was washed in PBS and bound protein was eluted with 0.1M glycine pH 2.5. Elution fractions contain pure IgG of the expected 150 kDa molecular weight (Figure 3.4).

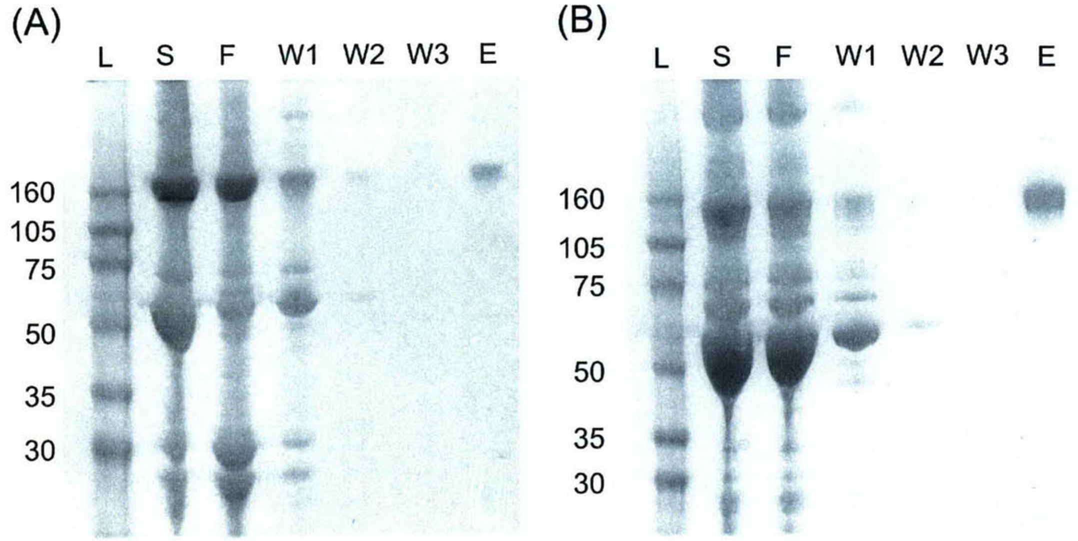


Figure 3.4. *Affinity purification.* Goat (A) or rabbit (B) serum was purified on an affinity column composed of streptavidin-agarose and biotinylated Fn3 (gI2.5.3T88I or rI4.5.5K27S/K56S). The serum (S), flowthrough (F), washes (W1-W3), and elution (E) were separated by polyacrylamide gel electrophoresis and stained with SimplyBlue SafeStain. (L) indicates the protein ladder.

Utility as Detection Reagents

The fluorophore DyLight633 was conjugated to amines on gI2.5.3T88I and rI4.5.5K27S/K56S. These fluorophore-Fn3 conjugates were used as secondary reagents in flow cytometry. Yeast displaying HA—Fn3—c-myc (irrelevant Fn3 clone) were labeled with anti-HA goat IgG followed by gI2.5.3T88I-DyLight633 and analyzed by flow cytometry. The cells displaying HA—Fn3—c-myc are clearly labeled whereas cells that lost plasmid (see *Materials and Methods*) or cells without primary antibody have only background signal (Figure 3.5A). Likewise, rI4.5.5K27S/K56S-DyLight633 effectively labels yeast initially labeled by anti-c-myc rabbit IgG (Figure 3.5B).

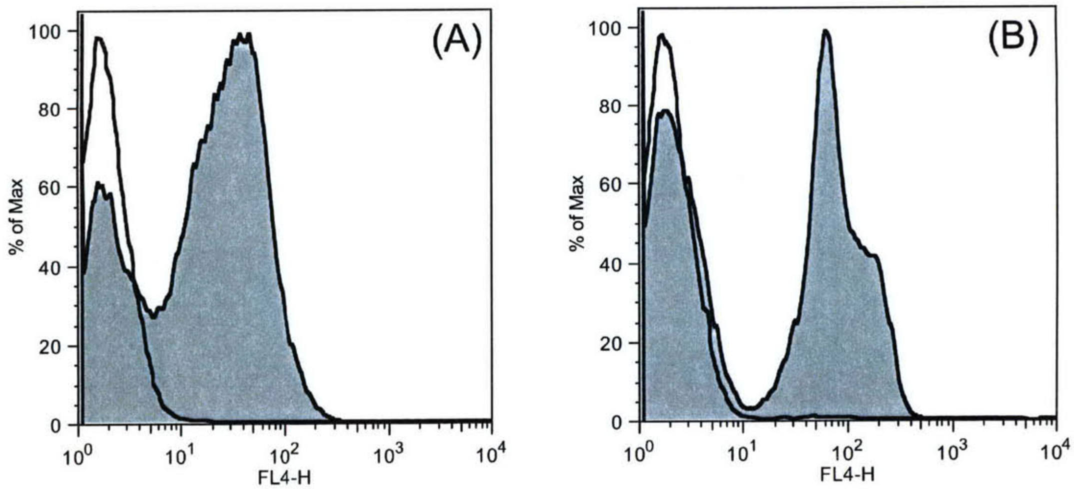


Figure 3.5. *Flow cytometry detection.* (A) Yeast displaying HA—Fn3—c-myc (irrelevant Fn3 clone) were labeled with PBS (empty) or anti-HA goat IgG (shaded) followed by DyLight633-conjugated gI2.5.3T88I and analyzed by flow cytometry. (B) as in (A) except anti-c-myc rabbit IgG and rI4.5.5K27S/K56S were used. Note that the two peaks in the IgG-labeled samples correspond to cells with and without plasmid, which serves as an effective internal control.

Discussion

Library creation and screenable throughput limit the number of clones that can be analyzed in a combinatorial library. Thus, the aim of library design is to enable sufficient shape and chemical diversity, within this limited sequence space, to provide high affinity binding to any desired epitope. Full amino acid diversity enables better shape and chemical complementarity but the vastness of sequence space may include sufficient nonfunctional sequences such as to reduce the overall frequency of binders in the screened library. Conversely, well-designed reduced diversity can improve the frequency of binders, but may not be able to mediate an interaction of equally high affinity. The current work provides direct competition of minimal and maximal diversity libraries in the fibronectin scaffold for molecular recognition of immunoglobulin G. Multiple binders of picomolar affinity for rabbit IgG and multiple binders with nanomolar affinity for goat IgG were isolated with a dominant preference for clones from the full diversity library. It is important to note that this comparison includes an effective dual mutagenesis approach for the evolution of lead clones enabling a more efficient broad search of sequence space. Though a serine/tyrosine binary code is effective in isolating binders of mid-nanomolar affinity, [3; 4] extensive diversity is more effective at generating low nanomolar to picomolar binding. Nevertheless, it remains true that amino acids have varied functionality and some, in particular tyrosine, are better suited for molecular recognition. Thus, the current results support the growing opinion that the ideal synthetic library design is a hybrid of full diversity and a binary serine/tyrosine code in which tyrosine is present at an increased level but more extensive alternatives than serine are included.[13]

The relative lack of library dominance in the DE loop (65% full diversity, 23% wild-type bias, 13% ambivalent) supports the hypothesis that the serine/tyrosine versus full diversity comparison was not largely affected by differences in the DE loop. In fact, given the prevalence of full diversity BC and FG loops in the selected binders, the increased presence of wild-type biased DE loops from the other library suggests that a wild-type bias is superior to random full diversity in this loop. This result is expected given its position on the edge of the diversified region as well as the ability of previously engineered Fn3 domains to bind with fully wild-type DE loops as previously discussed.[9]

The selected clones also demonstrate the importance of loop length diversity. BC loops of 6, 7, and 8 amino acids are observed multiple times. DE loops of wild-type length and one amino acid shorter occur. FG loops of all possible lengths except wild-type are observed. It is noteworthy that the longest allowed lengths in each loop were not observed in the sequenced binders. Longer loops are observed in nature as well as previously published binders and, thus, are tolerable in the scaffold. However, one could speculate that longer loops could be less structured resulting in entropic penalty upon binding. This phenomenon will require further monitoring as the collection of engineered fibronectin domains increases.

The efficacy of the engineered Fn3 domains in purification and detection further exemplifies the utility of this alternative scaffold for biotechnology applications and provides two useful high affinity reagents. The small size, single-domain architecture, and lack of disulfide bonds in Fn3 domains provide potential benefits over antibody-based reagents. Moreover, their facile production in bacteria and ease of site-specific labeling through amines or thiols simplify their use.

Materials and Methods

Fn3 Library Construction

Oligonucleotides were purchased from MWG Biotech and IDT DNA Technologies. The NNB library was previously constructed as described.[9] In the NNB library, the oligonucleotides encoding the BC, DE, and FG loops were replaced by NNB codons; specifically, the DNA encoding amino acids 23-30 (DAPAVTVR), 52-56 (GSKST), and 77-86 (GRGDSPASSK) were replaced with $(NNB)_x$ where $x = 6, 7, 8$, or 9 for the BC loop, $x = 4, 5, 6$, or 7 for the DE loop and $x = 5, 6, 8$, or 10 for the FG loop. The YS library replaced the DNA encoding amino acids 23-31 (DAPAVTVRY) and 77-86 with TMY codons that encode for serine and tyrosine in equal frequency. Loop lengths of 7, 8, 9, or 10 in the BC loop and 6, 7, 8 or 10 in the FG loop were included. Earlier selections from the NNB library rarely yielded FG loops of five amino acids, thus the YS library was constructed with a seven amino acid option rather than five. The DNA encoding amino acids 52-56 was replaced by a set of biased codons designed to yield 50% wild-type amino acid at G52, S53, S55, and T56. K54, because of its potential for steric and electrostatic hindrance of binding, was replaced by NNB_x where $x = 0, 1$, or 3. The degenerate portion of the oligonucleotide was ggBtcBNNBtcBacB where *a, c, g,* and *t* represent mixtures of 70% of the indicated nucleotide and 10% of each of the other three.

Full Fn3 genes were constructed by sequential annealing and extension of eight overlapping oligonucleotides. A 50 μL reaction was prepared with 0.2 μM oligonucleotide A (a2, b3, c6, or d7), 0.4 μM oligonucleotide B (a1, b4, c5, or d8), 1x polymerase buffer, 0.2 mM deoxynucleotide triphosphates, 1 mM $MgSO_4$, 1U KOD HotStart DNA Polymerase, 1M betaine, and 3% dimethyl sulfoxide. The mixture was denatured at 95° for 2 min., cycled ten times through 94° for 30 s, 58° for 30 s, and 68° for 1 min., and finally extended at 68° for 10 min. Forty microliters of products (a1+a2, b3+b4, c5+c6, d7+d8) were combined and thermally cycled at identical conditions. The appropriate strand (sense for a1+a2+b3+b4 and anti-sense for c5+c6+d7+d8) was amplified with 0.4 μM primer (p1 or p8) in a 100 μL reaction. The products were combined and thermally cycled under identical conditions. Full Fn3 genes were purified on an agarose gel, amplified by p1 and p8 in 100 μL reactions, and concentrated with PelletPaint. The plasmid acceptor vector pCTf1f4[14] was digested with NcoI, NdeI, and SmaI. Multiple aliquots of ~10 μg of Fn3 gene and ~3 μg of plasmid vector were combined with 50-100 μL of electrocompetent EBY100 and electroporated at 0.54 kV and 25 μF. Homologous recombination of the linearized vector and degenerate insert yielded intact plasmid. Cells were grown in YPD (10 g/L yeast extract, 20 g/L peptone, 20 g/L glucose) for 1 h at 30°, 250 rpm. The total number of transformants was determined by serial dilution plating on SD-CAA plates (0.1M sodium phosphate, pH 6.0, 182 g/L sorbitol, 6.7 g/L yeast nitrogen base, 5 g/L casamino acids, 20 g/L glucose). The library was propogated in SD-CAA, pH 5.3 (0.07M sodium citrate pH 5.3, 6.7 g/L yeast nitrogen base, 5 g/L casamino acids, 20 g/L glucose, 0.1 g/L kanamycin, 100 kU/L penicillin, and 0.1 g/L streptomycin) at 30°, 250 rpm.

Binder Selection and Affinity Maturation

Yeast were grown in SD-CAA at 30°, 250 rpm to logarithmic phase, pelleted, and resuspended to 1×10^7 cells/mL in SG-CAA (0.1M sodium phosphate, pH 6.0, 6.7 g/L yeast nitrogen base, 5 g/L casamino acids, 19 g/L galactose, 1 g/L glucose, 0.1 g/L kanamycin, 100 kU/L penicillin, and 0.1 g/L streptomycin) to induce protein expression. Induced cells were grown at 30°, 250 rpm for 8-24h.

Magnetic bead sorts consisted of a negative selection for clones that do not bind streptavidin-coated beads followed by a positive selection for clones that bind biotinylated IgG complexed to streptavidin-coated beads as described [15]. 0.75 μg of biotinylated goat or rabbit IgG (Rockland Immunochemicals, Gilbertsville, PA) was added to 4×10^6 streptavidin-coated magnetic Dynabeads (Invitrogen) in 1 mL PBSA (0.01 M sodium phosphate, pH 7.4, 0.137 M NaCl, 1 g/L bovine serum albumin) and incubated at 4° for 12-24 h. Beads were washed using a Dynal magnet with PBSA. Yeast displaying Fn3 were washed, resuspended in PBSA, and incubated with 4×10^6 IgG-free streptavidin beads for 2-12 h at 4°. A magnet was applied to the cell/bead mixture and unbound cells were collected. The washed IgG-labeled beads were added to these cells and incubated at 4° for 2-12h. The beads were applied to the magnet and washed with PBSA. The beads and attached cells were transferred to SD-CAA for growth.

The naïve library was sorted twice (zero washes at 4°, one wash at 4°) with growth and induction after each sort. The resultant population was labeled with 150 nM mouse anti-c-myc antibody (clone 9E10) followed by 25 nM goat anti-mouse phycoerythrin conjugate. Full-length Fn3 clones, represented by cells with a positive phycoerythrin signal, were selected via FACS. Plasmid DNA was extracted and mutagenized as described [9]. Error-prone PCR was performed on the full gene and each of the three loops; the mutated gene or shuffled combinations of the mutated loops were co-transformed with linearized plasmid vector to produce intact plasmid via homologous recombination. Transformed yeast were grown in SD-CAA for further selection. The mutagenized population was sorted twice on magnetic beads (one wash at 4°, one wash at 22°) followed by c-myc$^+$ FACS and further mutagenesis. After two magnetic bead sorts (one wash at 22°, two washes at 22°), binding to soluble IgG was assayed by flow cytometry. Yeast were labeled with 3.3 nM biotinylated IgG followed by 33 nM streptavidin-phycoerythrin conjugate. Cells with the highest phycoerythrin signal were collected by FACS and mutated. Remaining selections were performed with 20-500 pM biotinylated IgG and 67 nM chicken anti-c-myc followed by 150 nM streptavidin-fluorophore and 25 nM bovine anti-chicken phycoerythrin conjugate. Cells with the highest fluorophore:phycoerythrin ratio were selected by FACS.

DNA Sequencing and Point Mutations

Multiple clones from several populations were sequenced. Plasmid DNA was isolated using the Zymoprep kit II, cleaned using the Qiagen PCR Purification kit, and transformed into DH5α or XL1-Blue *E. coli*. Individual clones were grown, miniprepped, and sequenced using BigDye chemistry on an Applied Biosystems 3730. Single amino acid mutations were introduced by standard site-directed mutagenesis using the QuikChange Mutagenesis Kit according to the manufacturer's instructions. Clone construction was verified by DNA sequencing.

Affinity Measurement

The plasmid for the clone of interest was transformed into yeast using the Frozen EZ Transformation Kit II, and cells were grown and induced as for selection. Cells were washed in PBSA and resuspended in PBSA containing fluorophore-conjugated IgG over a range of concentrations. Sample volumes and cell densities were selected to ensure tenfold excess of IgG relative to displayed Fn3. Samples were incubated at 22° for a sufficient time to ensure the approach to equilibrium was at least 98% complete. After incubation, cells were washed and analyzed on a FACS Calibur cytometer (Becton Dickinson). The relative binding was calculated by subtracting background signal, which was determined in an unlabeled control, and normalizing to the saturated signal at high concentrations. The equilibrium dissociation constant, K_d, was identified as the concentration corresponding to half-maximal binding.

Stability

The yeast surface display thermal denaturation assay[12] was performed as described[9]. Yeast displaying the clone of interest were washed and resuspended in PBSA, incubated at 22-85° for 30 min., and incubated on ice for 5 min. The cells were incubated in 20 nM fluorescein-conjugated IgG for 30 min., washed, and analyzed on an Epics XL flow cytometer. The minimum and maximum fluorescence, the midpoint of thermal denaturation ($T_{1/2}$), and the enthalpy of unfolding at $T_{1/2}$ were determined by minimizing the sum of squared errors between experimental data and theoretical values according to a two-state unfolding equation.

Specificity

Yeast displaying the clone of interest were incubated with 100 nM bovine IgG-phycoerythrin (Santa Cruz Biotechnology, Santa Cruz, CA), AlexaFluor488 conjugates of streptavidin, chicken IgG or mouse IgG (Invitrogen), or fluorescein conjugates of goat IgG, human IgG, or rabbit IgG (Sigma). Cells were incubated for 30 min., washed with

PBSA, and analyzed by flow cytometry. To test lysozyme binding, cells were incubated with 100 nM biotinylated lysozyme for 30 min., washed, and resuspended in 25 nM AlexaFluor488-conjugated streptavidin. Cells were washed and analyzed by flow cytometry. Fluorophore signal was compared to both unlabeled and non-displaying cells.

Fn3 Production

The Fn3 gene was digested with NheI and BamHI and transformed to a pET vector containing a HHHHHHKGSGK-encoding C-terminus. The six histidines enable metal affinity purification, and the pentapeptide provides two additional amines for chemical conjugation. The plasmid was transformed into Rosetta (DE3) *E. coli* (Novagen), which was grown in LB medium with 100 mg/L kanamycin and 34 mg/L chloramphenicol at 37°. Two hundred μL of overnight culture was added to 100 mL of LB medium, grown to an optical density of 0.5 units, and induced with 0.5 mM IPTG overnight. Cells were pelleted, resuspended in lysis buffer (50 mM sodium phosphate, pH 8.0, 0.5M NaCl, 5% glycerol, 5 mM CHAPS, 25 mM imidazole, and 1x complete EDTA-free protease inhibitor cocktail (Roche, Indianapolis, IN)), and exposed to four freeze-thaw cycles. The soluble fraction was clarified by centrifugation at 15,000*g* for 10 min. and purified by metal affinity chromatography on TALON resin.

Affinity Purification

Purified Fn3 (gI2.5.3T88I and rI4.5.5K27S/K56S) was biotinylated using EZ-Link Sulfo-N-hydroxysuccinimide-LC-biotin (Pierce, Rockford, IL) according to the manufacturer's instructions. Excess biotin was removed using a Zeba desalting spin column (Pierce). Biotinylated Fn3 was added to 1 mL of strepavidin-agarose (Pierce) in a column and washed. Goat or rabbit serum was added to the column and flowthrough was reapplied once. The column was washed with three 5 mL aliquots of PBS. Protein was eluted with 0.1M glycine, pH 2.5. The original serum, flowthrough, washes, and elution were separated by SDS-PAGE on 12% BisTris gel (Invitrogen) in the absence of dithiothreitol. The gel was stained with SimplyBlue SafeStain (Invitrogen) and imaged.

Detection

Purified Fn3 (gI2.5.3T88I and rI4.5.5K27S/K56S) was labeled with DyLight633 N-hydroxysuccinimide-ester (Pierce) according to the manufacturer's instructions. Unreacted dye was removed using a Zeba desalting spin column. Yeast were induced to display an irrelevant Fn3 clone with the HA and c-myc epitopes. As in all yeast surface display, a fraction of this population does not display any Fn3 as a result of plasmid loss. These cells serve as an internal negative control. One million yeast were incubated with 50 nM anti-HA goat IgG or anti-c-myc rabbit IgG (Genscript, Piscataway, NJ), washed, and incubated with 50 nM DyLight633-conjugated Fn3. Cells were washed and analyzed on a FACS Calibur cytometer.

References

1. Fellouse, F., Barthelemy, P. A., Kelley, R. F. & Sidhu, S. (2006). Tyrosine plays a dominant functional role in the paratope of a synthetic antibody derived from a four amino acid code. *Journal of Molecular Biology* **357**, 100-14.
2. Fellouse, F., Wiesmann, C. & Sidhu, S. (2004). Synthetic antibodies from a four-amino-acid code: a dominant role for tyrosine in antigen recognition. *Proc Natl Acad Sci USA* **101**, 12467-72.
3. Fellouse, F., Li, B., Compaan, D. M., Peden, A. A., Hymowitz, S. G. & Sidhu, S. (2005). Molecular recognition by a binary code. *Journal of Molecular Biology* **348**, 1153-62.
4. Koide, A., Gilbreth, R. N., Esaki, K., Tereshko, V. & Koide, S. (2007). High-affinity single-domain binding proteins with a binary-code interface. *Proc Natl Acad Sci USA* **104**, 6632-7.
5. Fellouse, F., Esaki, K., Birtalan, S., Raptis, D., Cancasci, V. J., Koide, A., Jhurani, P., Vasser, M., Wiesmann, C., Kossiakoff, A. A., Koide, S. & Sidhu, S. (2007). High-throughput generation of synthetic antibodies from highly functional minimalist phage-displayed libraries. *Journal of Molecular Biology* **373**, 924-40.
6. Richards, J., Miller, M., Abend, J., Koide, A., Koide, S. & Dewhurst, S. (2003). Engineered fibronectin type III domain with a RGDWXE sequence binds with enhanced affinity and specificity to human alphavbeta3 integrin. *Journal of Molecular Biology* **326**, 1475-88.
7. Karatan, E., Merguerian, M., Han, Z., Scholle, M. D., Koide, S. & Kay, B. K. (2004). Molecular recognition properties of FN3 monobodies that bind the Src SH3 domain. *Chemistry & Biology* **11**, 835-44.
8. Huang, J., Koide, A., Nettle, K., Greene, G. & Koide, S. (2006). Conformation-specific affinity purification of proteins using engineered binding proteins: Application to the estrogen receptor. *Protein Expression and Purification* **47**, 348-354.
9. Hackel, B., Kapila, A. & Wittrup, K. (2008). Picomolar affinity fibronectin domains engineered utilizing loop length diversity, recursive mutagenesis, and loop shuffling. *Journal of Molecular Biology* **381**, 1238-52.
10. Shusta, E. V., Kieke, M. C., Parke, E., Kranz, D. M. & Wittrup, K. D. (1999). Yeast polypeptide fusion surface display levels predict thermal stability and soluble secretion efficiency. *Journal of Molecular Biology* **292**, 949-56.
11. Main, A. L., Harvey, T. S., Baron, M., Boyd, J. & Campbell, I. D. (1992). The three-dimensional structure of the tenth type III module of fibronectin: an insight into RGD-mediated interactions. *Cell* **71**, 671-8.
12. Orr, B. A., Carr, L. M., Wittrup, K., Roy, E. J. & Kranz, D. M. (2003). Rapid method for measuring ScFv thermal stability by yeast surface display. *Biotechnol Prog* **19**, 631-8.

13. Koide, S. & Sidhu, S. S. (2009). The Importance of Being Tyrosine: Lessons in Molecular Recognition from Minimalist Synthetic Binding Proteins. *ACS Chem Biol.*
14. Lipovsek, D., Lippow, S., Hackel, B., Gregson, M. W., Cheng, P., Kapila, A. & Wittrup, K. (2007). Evolution of an interloop disulfide bond in high-affinity antibody mimics based on fibronectin type III domain and selected by yeast surface display: molecular convergence with single-domain camelid and shark antibodies. *Journal of Molecular Biology* **368**, 1024-41.
15. Ackerman, M., Levary, D., Tobon, G., Hackel, B., Orcutt, K. D. & Wittrup, K. D. (2009). Highly avid magnetic bead capture: An efficient selection method for de novo protein engineering utilizing yeast surface display. *Biotechnol Prog.*

4. STABILITY AND COMPLEMENTARITY BIAS IMPROVE THE PROTEIN FUNCTIONALITY LANDSCAPE

Introduction

The design and construction of synthetic combinatorial libraries is critical for the development of alternative scaffolds for molecular recognition[1] as well as for high throughput approaches to antibody engineering such as those required for proteomic applications.[2] Design requires the balance of creating sufficient conformational and chemical diversity to yield high affinity binding to myriad epitopes while providing a high frequency of functional clones such that the limited searchable sequence space contains clones with the appropriate phenotype. Continued study of library design and construction will enable more efficient selection of high affinity binders from a variety of scaffolds including Fn3. We sought to develop an improved Fn3 library design through incorporation of two key features: wild-type conservation of residues that are structurally critical and/or are less likely to contribute to the desired binding interaction and tailored amino acid diversity biased to functional amino acids.

Despite their location in the BC/DE/FG loop region of Fn3, some residues may be critical to the conformational stability of the protein fold. As such, diversification of these positions may produce a library population with reduced average stability. Destabilization limits the robustness of binders in biotechnology applications such as the stringent washing steps of purification and detection. Instability can result in degradation and aggregation of *in vivo* diagnostics and therapeutics, which reduces potency and can elicit an immune response. Moreover, destabilization decreases the tolerance to mutation, which decreases the capacity for evolution.[3] Also, the potentially resultant flexibility may diminish the free energy change upon binding because of entropic effects. Moreover, conservation at structurally critical positions enables diversity to be focused on positions that are more likely to contribute to the binding interaction yielding a more efficient search of sequence space. In the current work, we use stability, structural, and sequence analyses to identify conservation sites that may benefit library design.

Early library designs commonly used NNB or NNS/NNK randomized codons to approximate an equal distribution of all amino acids. Yet not all amino acids are equivalent in their ability to provide conformational and chemical complementarity for molecular recognition so a tailored distribution may be more effective. Sidhu and colleagues have investigated this hypothesis and demonstrated the utility of a tyrosine/serine library as well as the unique efficacy of tyrosine to mediate molecular recognition in antibody fragments.[4-6] Direct competition of full diversity and tyrosine/serine diversity libraries in the Fn3 domain was dominated by the full diversity library for selection of high affinity binders to goat and rabbit immunoglobulin G.[7] Thus, though tyrosine/serine may provide ample diversity for binding, an expanded repertoire enables higher complementarity. The expanded repertoire can be effectively utilized with an efficient library design and/or affinity maturation scheme. A tailored antibody library with elevated tyrosine, glycine, and serine and low levels of all other amino acids except cysteine was superior to a tyrosine/serine library.[8] A similarly biased library was used with the Fn3 scaffold to yield a 6 nM binder to maltose binding protein[9] and a novel 'affinity clamp' for peptide recognition.[10] These biased distributions were created by oligonucleotide synthesis using custom trimer phosphoramidite mixtures. The current work investigates the ability to create a desired distribution via inexpensive skewed nucleotide mixtures. In particular, the amino acid distribution in human and mouse CDR-H3 loops is effectively mimicked. We demonstrate that a new library incorporating selective conservation and tailored diversity is superior to both an unbiased library with approximately equal amino acid diversity and a tyrosine/serine binary code library. This library enabled the generation of binders to a multitude of targets with potential utility in research, biotechnology, and therapy.

Results

Fn3 Stability

We used yeast surface display for efficient stability analysis of Fn3 clones. It has been demonstrated that the number of displayed single-chain T-cell receptors per yeast cell correlates to receptor stability.[11] To validate this correlation for Fn3, we created yeast surface display vectors of binders to vascular endothelial growth factor receptor 2 spanning a range of stabilities: free energies of unfolding from 3.8 to 7.5 kcal/mol and midpoints of thermal denaturation of 42 to 84°.[12] Clonal cultures of yeast were grown at 30°, Fn3 expression was induced at 37°, and the amount of displayed Fn3 was quantified by flow cytometry. The clones exhibit a positive relationship between display and stability spanning a substantial display range between the least and most stable clones (Figure 4.1A) thereby validating this technique for stability comparison.

This validated approach was used to explore domain stabilization via single-site wild-type conservation in the context of a diverse library. To quantify this impact, a series of libraries were constructed: one library with fully diversified BC, DE, and FG loops and multiple libraries of the same design except for wild-type conservation at a single position of interest. The libraries were transformed into a yeast surface display system and the amount of Fn3 displayed upon induction at 37° was quantified by flow cytometry. Eleven of fourteen positions studied, as well as a multisite library, exhibit improved display with wild-type conservation. A26, V27, and T28 have increased display but not of statistical significance (Figure 4.1B).

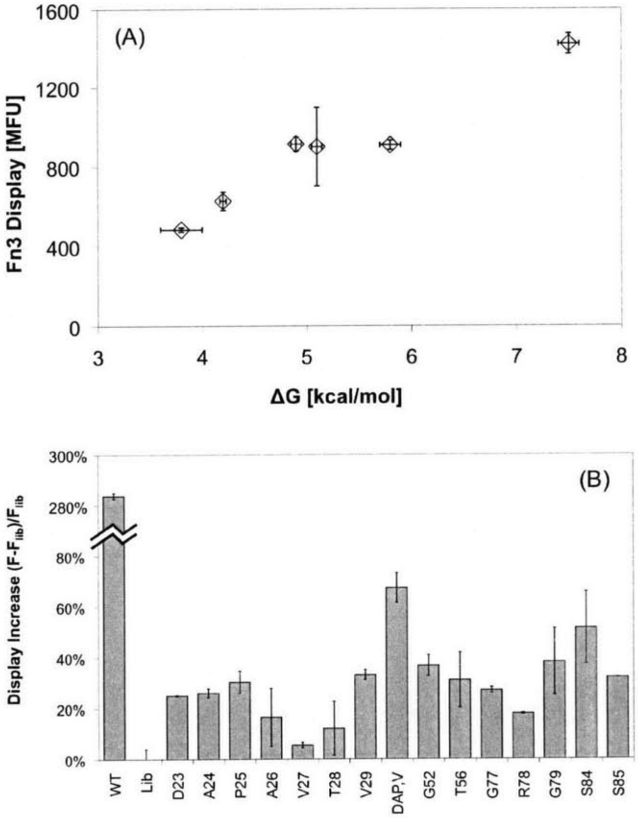


Figure 4.1. *Fn3 stability.* Yeast clones or libraries were grown to logarithmic growth phase at 30°. Expression of Aga2p-Fn3 was induced at 37°. Fn3 display level was quantified by flow cytometry using mouse αc-myc antibody and αmouse antibody-R-phycoerythrin conjugate. (A) The extent of Fn3 display of VEGF-R2 binders was determined and compared to the previously determined stability.[12] (B) Libraries were created with full diversity in the BC, DE, and FG loops except maintenance of wild-type at the position indicated. The display level of this singly-constrained library was compared to a non-constrained library.

Solvent Accessible Surface Area

The solvent accessible surface area of each potentially diversified position was calculated using GetArea[13] for wild-type Fn3 (solution structure 1TTG[14] and crystal structures 1FNA[15]) and an engineered binder (2OBG[16]). Despite their presence in previously diversified loop regions, the side chains of D23, A24, P25, V29, G52, and S85 are relatively inaccessible; peripheral residues W22, Y32, A57, T76, and P87 are also buried (Figure 4.2). Conversely, the amino acids in the middle of each loop are relatively exposed, supporting the ability of these sites to be diversified while maintaining the correct fold.

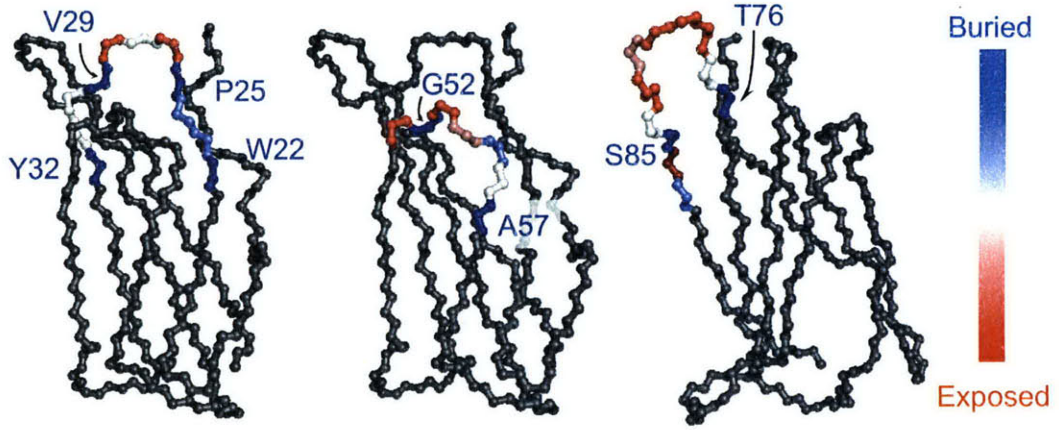


Figure 4.2. *Solvent accessibility.* The solvent accessible surface area (SASA) was calculated for each residue using GetArea[13] with a 1.4Å probe. The SASA of each side chain in the solution (1TTG[14]) and crystal (1FNA[15]) structures of wild-type Fn3 and an engineered binder (2OBG[16]) were normalized by the SASA of the side chain in a random coiled peptide. The fibronectin domain is presented using 1TTG with the residues in the loop regions (W22-Y32, P51-A57, T76-P87) color-coded according to the mean value of the accessibility ratios for the three cases.

Sequence Analysis

The mutational flexibility of each position was further explored through phylogenetic sequence analysis. The type III domains of fibronectin in chimpanzee, cow, dog, horse, human, mouse, opossum, platypus, rat, and rhesus monkey were aligned, and the relative frequency of each amino acid was determined (Figure 4.3A). The peripheral residues W22, Y32, P51, A57, and P87 are well conserved; however, T76 is variable. Other sites exhibiting conservation three-fold above random are A24 (22%), P25 (62%), V29 (25% as well as 43% isoleucine), G52 (25%), S53 (23%), S55 (27%), G77 (21%), G79 (19%), and S85 (66%); also note that T56 is 12% conserved with 51% of the homolog serine. Thus, the BC loop exhibits conservation of its peripheral hydrophobic residues except Y31. The DE loop, except for the central lysine, is well-conserved. The FG loop has a trend towards glycine from G77 to G79 and two highly conserved sites near the C-terminus.

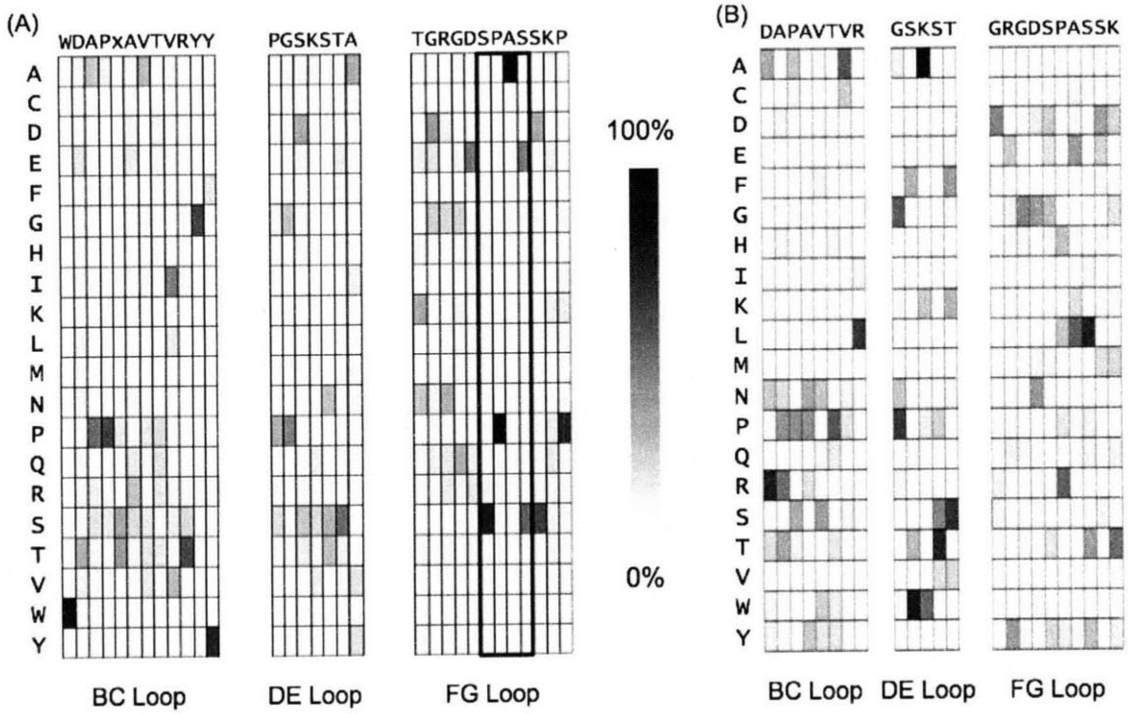


Figure 4.3. *Sequence analysis.* Amino acid sequences were aligned. The amino acid frequency at each position is presented in an intensity scale in each column. The wild-type human sequence for the tenth type III domain is presented at the top of each column (W22-Y32, P51-A57, T76-P87). The *x* in the BC loop corresponds to an amino acid present in other domains that is not present in the human tenth type III domain. (A) Wild-type analysis. The amino acid sequences for the type III domains of fibronectin in chimpanzee, cow, dog, horse, human, mouse, opossum, platypus, rat, and rhesus monkey were analyzed. The outline around S81-S84 represents rare positions as most type III domains contain shorter FG loops. (B) Binder sequence analysis. The amino acid sequences for binder-engineered Fn3 domains were aligned. The amino acid frequency at each position was compared to the frequency in the composite naïve libraries.

Published sequences of engineered binders were analyzed similarly; though in this analysis amino acid frequencies must be compared to expected frequencies based on variable library designs (Figure 4.3B). Wild-type is present at least twice as often in binders as in the naïve library at three positions: P25 (15% in binders *versus* 5% in libraries), G52 (26% *v.* 13%), and G79 (17% *v.* 5%). In addition, three positions yield substantial enrichment of homologs: alanine at V29 (20% *v.* 6%), threonine at S55 (25% *v.* 6%), and serine at T56 (28% *v.* 11%).

Library Design

The stability, accessibility, and sequence analyses (summarized in Table 4.1) were used to determine the degree of diversification desired at each position. For example, proline at position 25 significantly stabilizes the library, is essentially inaccessible to solvent, and is highly conserved in the type III fibronectin domains of mammals. Thus, the new library will be heavily biased towards proline at this position. Conversely, the adjacent alanine at position 26 does not significantly stabilize the library, is highly accessible, and exhibits essentially no conservation. As a result, this position will be fully diversified in the new library design.

Along with conservation bias to maintain structural integrity and focus diversity on positions better suited for molecular recognition, it was desired to bias the diversity to functional amino acids. Tyrosine has demonstrated unique utility in molecular recognition.[4-6] Glycine provides conformational flexibility. Serine and alanine are valuable as small, neutral side chains. Acidic residues, arginine, and lysine provide charge although the utility is unclear.[17] Other side chains may provide ideal complementarity in less frequent situations. Thus, we propose the ideal diversity contains high tyrosine, glycine, and serine and/or alanine as well as small levels of all other amino acids. For the particular amino acid distribution we sought guidance from natural molecular recognition. The amino acid distribution in CDR-H3 matches the desired diversity and was used as the library design model (Figure 4.4). Each position was designed to incorporate the desired level of wild-type conservation and to match the antibody CDR-H3 repertoire in the non-conserved portion of the distribution. The DE loop is a slight exception because a very similar design was previously validated as effective.[7] In this loop, G52, S53, S55, and T56 are highly conserved with wild-type at 50% frequency and unbiased distribution of all other amino acids. The lack of antibody-inspired bias in this loop is of limited detriment because of the high conservation of the wild-type amino acids. Multiple loop lengths, selected based on phylogenetic occurrence,[18] are included in each loop. The resultant library design is summarized in Table 4.1.

Table 4.1. *Fn3 library design summary. Pos.* and *WT* are the amino acid position and residue in the human wild-type tenth type III domain. *Access.* is the ratio of solvent accessible surface area for the residue in the fibronectin domain compared to the residue in a random coiled peptide. *Stability* is the relative increase in yeast surface display level of a library with wild-type conservation at the position of interest. *Native* indicates the frequencies of the indicated amino acids in type III fibronectin domains of ten species. *Binders* indicates the enrichment of wild-type (or homolog as indicated) in engineered binders relative to the naïve frequency. *Library Design* indicates the intended amino acid distribution in the new library. *Ab div.* is the designed amino acid distribution that mimics antibody CDR-H3. * indicates the location of loop length variability.

				Sequences		
Pos.	WT	Access.	Stability	Native	Binders	Library Design
				BC Loop		
22	W	1	-	100% W	-	wild-type
23	D	35	25 ± 0%	5% D + 14% E	1.0x	Ab div. (10% D)
24	A	32	26 ± 2%	22% A + 12% S	0.4x	8% A + Ab div.
25	P	10	30 ± 4%	62% P	2.7x	42% P + Ab div.
26	A	75	16 ± 11%	7% A	0.7x	Ab div. *
27	V	57	5 ± 1%	14% V	0.4x	Ab div. *
28	T	75	12 ± 11%	12% T	1.6x	Ab div. *
29	V	3	33 ± 2%	25% V + 56% I+L	0.6x (3.2x A)	25% A|L|S|V
30	R	49	-	6% R	1.1x	Ab div.
31	Y	43	-	6% Y, 62% G	-	50% S, 50% Y
32	Y	1	-	75% Y	-	wild-type
				DE Loop		
51	P	79	-	31% P	-	wild-type
52	G	12	37 ± 4%	25% G	2.0x	49% G
53	S	83	-	23% S + 14% T	1.0x	50% S
54	K	64	-	6% K	3.6x	NNB div. *
55	S	41	-	27% S + 26% T	1.6x (4.2x T)	50% S
56	T	48	31 ± 11%	12% T + 51% S	0.8x (2.6x S)	49% T
57	A	8	-	32% A	-	wild-type
				FG Loop		
76	T	8	-	7% T	-	wild-type
77	G	48	27 ± 1%	21% G	0.7x	12% G + Ab div.
78	R	81	18 ± 0%	12% R	1.5x	Ab div.
79	G	77	38 ± 13%	19% G	2.7x	12% G + Ab div.
80	D	74	-	7% D + 48% E	1.9x	Ab div. *
81	S	69	-	rare	0.7x	Ab div. *
82	P	76	-	rare	1.5x	Ab div. *
83	A	81	-	rare	0.6x	Ab div. *
84	S	54	52 ± 14%	rare	0.5x	Ab div. *
85	S	14	32 ± 0%	66% S	1.2x	100% S
86	K	88	-	12% K	1.5x	Ab div.
87	P	40	-	74% P	-	wild-type

Library Construction

Though trimer phosphoramidite library construction enables precise creation of unique amino acid distributions, this approach is expensive with the inclusion of multiple specialty codon mixtures. As an inexpensive alternative, standard oligonucleotide synthesis was employed using custom mixtures of skewed nucleotides at each position. The optimal set of three nucleotide mixtures was determined for each codon as follows. All possible sets of nucleotide mixtures with each component at 5% increments were filtered to select only those that closely match the desired levels of wild-type and tyrosine and reasonably match glycine, serine, aspartic acid, alanine, and arginine; these amino acids are the most frequent in antibody CDR-H3 and are functionally diverse. Sample protein libraries were then produced *in silico* from the amino acid probability distributions resulting from the sets of nucleotide mixtures. The library calculated to be most likely to be produced from the intended distribution (*i.e.*, the antibody repertoire with the appropriate wild-type bias) was selected as optimal. This process was repeated for each position in the library. In general, these skewed nucleotide mixtures provide good matches to the desired amino acid distributions (Figure 4.4). The two exceptions are decreased levels of glycine and elevated cysteine. Since the latter two positions in a cysteine codon (TGT or TGC) are shared by glycine (GGN), it is not possible to create high levels of glycine without also yielding high cysteine unless TNN codons are depleted, which depletes tyrosine. Thus, a compromise is reached with 6% glycine and 10% cysteine. Though this incorporates a relatively high level of cysteine, the library design still yields many cysteine-free clones; moreover, interloop disulfide bonds are a potentially advantageous element.[19]

Fn3 genes were constructed by overlap extension PCR of partially degenerate oligonucleotides. Transformation into yeast by electroporation with homologous recombination yielded $2.5\text{x}10^8$ transformants. Sequencing and flow cytometry analysis indicate 60% of clones encode for full-length Fn3 resulting in $1.5\text{x}10^8$ Fn3 clones. Sequence analysis reveals that the skewed nucleotides accurately match their intended distribution (Figure 4.4). The library is termed G4, as it is the fourth generation Fn3 library created in our laboratory after the two-loop, single-length BF14 library[19], the three-loop, length-diversified NNB library[18], and the three-loop, DE-conserved tyrosine/serine library YS[7].

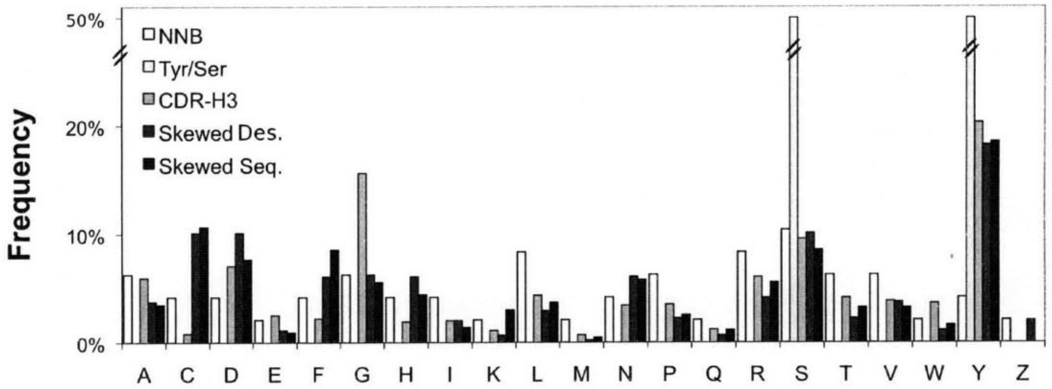


Figure 4.4. Amino acid distributions. The frequencies of each amino acid in multiple distributions are presented. *NNB* refers to a degenerate codon with 25% of each nucleotide at the first two positions and 33% of C, T, and G at the third position. *Tyr/Ser* refers to an even mix of tyrosine and serine. *CDR-H3* refers to the expressed human and mouse CDR-H3 sequences.[20] *Skewed Des.* refers to the theoretical distribution attainable using skewed oligonucleotides. *Skewed Seq.* refers to the distribution attained experimentally using skewed nucleotides.

Library Comparison

The new G4 library design was compared to a non-conserved, full diversity library (NNB[18]) and a library with wild-type conservation in the DE loop only and tyrosine/serine diversity (YS[7]) (Table 4.2). The libraries were pooled for comparison and tested for their ability to generate binders to seven targets: human A33, mouse A33, epidermal growth factor receptor (EGFR), Fcγ receptors IIA and IIIA (FcγRIIA and FcγRIIIA), mouse immunoglobulin G (mIgG), and human serum albumin (HSA). The naïve library was sorted by magnetic bead selections,[21] and lead clones were diversified by error-prone PCR on the full Fn3 gene and shuffling of mutagenized Fn3 loops. Multiple rounds of selection and diversification were performed to yield binders to each target. Sequence analysis of each binding population revealed that 19 of 21 binders originated from the G4 library while two clones were likely of NNB origin and no YS clones were identified (Table 4.3, Figure 4.5). Given the comparable number of clones in the naïve libraries, this result indicates that G4 is a superior library design to both NNB and YS for the selection of protein binders.

Table 4.2. Library design. *Loop Diversity* indicates the library of codons included at positions without wild-type bias. *Biased Positions* indicates positions within the diversified loops (23-31, 52-56, 77-86) that are biased towards wild-type. *Full-length Fn3s* indicates the library size; *i.e.,* the number of yeast transformants that encode for full-length Fn3 domains.

Library	Loop Diversity	Biased Positions	Full-length Fn3s
NNB	full diversity (NNB codons)	none	0.7×10^8
YS	50% Y, 50% S	52, 53, 55, 56	1.5×10^8
G4	antibody-based (18% Y, 10% S, ...)	23, 24, 25, 29, 31, 52, 53, 55, 56, 77, 79, 85	1.5×10^8

Table 4.3. Engineered binder sequences. *Name* is the name of each clone. *Target* is the cognate protein bound by the Fn3 clone. *23* refers to the amino acid present at position 23, which is aspartic acid (D) in wild-type Fn3; all positions diversified in the naïve library are likewise presented. *Framework* refers to amino acid mutations outside of the diversified loops. A dash indicates no amino acid.

Name	Target	23	24	25	26	27	28	-	29	30	31	52	53	54	55	56	77	78	79	80	81	82	83	84	85	86	Framework	K_d [nM]
WT	-	D	A	P	A	V	T	-	V	R	Y	G	S	K	S	T	G	R	G	D	S	P	A	S	S	K	-	-
E4.2.1	EGFR	Y	G	F	S	L	-	-	A	S	S	R	S	P	W	F	S	N	D	F	S	N	R	Y	S	G	-	0.25 ± 0.07
E6.2.6	EGFR	F	D	Y	A	-	-	-	V	T	Y	G	W	I	S	T	D	N	S	H	W	P	F	R	S	T	I90T	0.26 ± 0.13
E6.2.10	EGFR	Y	L	R	D	P	R	Y	V	D	Y	W	Y	L	P	E	Y	D	G	Y	R	E	S	T	P	L	-	0.96 ± 0.11
EI1.4.1	EGFR	Y	G	P	F	Y	Y	V	A	H	S	R	S	P	W	F	S	K	C	Y	D	G	-	-	S	V	-	0.85 ± 0.50
EI2.4.6	EGFR	Y	H	P	F	Y	Y	V	A	H	S	R	S	P	W	F	D	S	N	G	-	-	-	-	S	H	-	2.9 ± 0.3
EI3.4.2	EGFR	Y	G	S	S	Y	-	-	A	S	Y	R	S	P	W	F	P	S	G	I	-	-	-	-	S	A	T58I	9.5 ± 3.5
EI3.4.3	EGFR	L	H	H	R	S	D	-	V	R	S	G	S	R	S	L	W	G	S	Y	C	C	-	-	S	N	E47K	0.25 ± 0.05
EI4.4.2	EGFR	Y	F	R	D	P	R	Y	V	D	Y	W	Y	L	P	E	G	D	D	Q	N	A	-	-	G	L	V45A	1.4 ± 0.2
IIa8.2.6	FcγIIa	C	T	H	L	H	-	-	W	D	Y	A	L	C	P	G	V	G	G	D	-	-	-	-	D	W	R6G, T35F, V72A I88S, K98E	850
IIIa6.2.6	FcγIIIa	D	M	P	F	-	-	-	S	D	S	G	T	D	S	L	S	S	G	S	N	-	-	-	S	Y	A12V, S21N, T35A	530
hA2.2.1	hA33	Y	C	P	D	G	C	H	S	Y	Y	R	S	I	S	S	F	R	W	P	-	-	-	-	S	F	-	-
hA2.2.2	hA33	N	T	Y	F	S	F	-	L	Y	Y	S	S	L	H	T	G	T	W	P	-	-	-	-	S	Y	-	-
hA3.2.1	hA33	S	Y	S	S	Y	N	S	W	D	S	N	S	D	C	I	R	D	C	D	F	Y	-	-	S	Y	Y32F	-
hA3.2.2	hA33	Y	Y	H	L	R	G	-	L	D	S	R	S	Y	S	T	V	N	D	Y	I	-	-	-	S	Y	S21G Q46K T49A	-
mA3.2.1	mA33	S	S	S	L	Y	N	-	S	A	Y	V	W	D	C	T	P	N	Y	S	F	-	-	-	S	L	Y32F	-
mA3.2.2	mA33	C	C	L	F	F	-	-	S	G	Y	G	L	V	Y	W	D	N	V	G	-	-	-	-	S	N	I90V	-
mA3.2.3	mA33	S	F	P	C	V	-	-	S	S	S	G	D	T	T	S	S	T	C	Y	P	-	-	-	S	Y	-	-
mA3.2.4	mA33	S	C	P	I	C	P	R	A	T	S	A	T	-	S	S	D	Q	G	Y	D	D	-	-	S	A	I34V	-
mA3.2.5	mA33	Q	C	H	Y	Y	Y	-	A	Q	S	S	S	K	S	T	Y	N	W	F	L	D	S	V	S	I	A12V	-
Alb3.2.1	hAlb	G	A	P	A	C	-	-	A	A	Y	G	S	G	T	S	S	R	Y	Y	Y	C	-	-	S	E	-	-
mI2.2.1	mIgG	C	C	S	D	N	C	-	S	N	S	R	S	C	F	M	D	S	N	G	-	-	-	-	P	H	V72A	4.1 ± 0.7

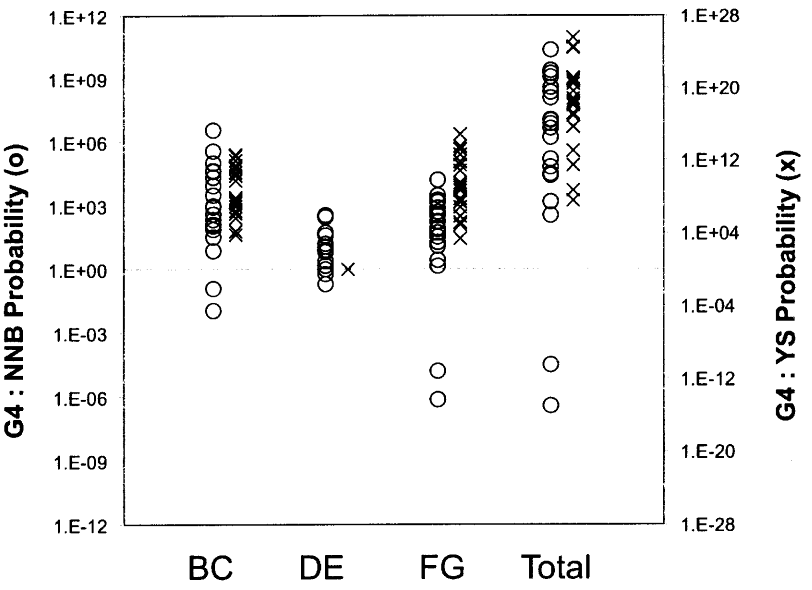


Figure 4.5. Library source probability. For each binding clone sequence, the probability of origination from each library was calculated based on library design. The relative preferences for G4 *versus* NNB (o) or G4 *versus* YS (x) are presented for each loop as well as the total domain. Each symbol indicates a sequenced clone.

Sequence analysis reveals that wild-type bias is approximately maintained or perhaps slightly reduced in the BC and FG loops of binders while the strong bias at G52, S55, and T56 is slightly reduced but still highly frequent (Figure 4.6a). It is noteworthy that in addition to 20% occurrence at G79, glycine is present at 15% at position 80. At position 29, equal amounts of alanine, leucine, serine, and wild-type valine were included in the naïve library; in binders, the smallest available side-chain, alanine, is present at 35% while the largest side-chain, leucine, occurs with only 10% frequency. Cumulative analysis of amino acid frequency at positions without wild-type bias indicates maintenance of the preferentially high levels of tyrosine, serine, glycine, aspartic acid, and arginine (Figure 4.6b). Conversely, cysteine and histidine, which were included at higher frequency than intended because of their codon similarity to tyrosine, are present at reduced levels in binders. Eight of nineteen (42%) G4-based binders are cysteine-free as compared to 19% in the naïve library. Interestingly, only three clones (16%) have a single cysteine as compared to a naïve 33% whereas seven clones (37%) contain two cysteines (26% in naïve library). A single clone has four cysteines. Thus, a strong selective pressure exists against unpaired cysteines. Of particular interest, six of the seven two-cysteine clones contain cysteine residues in identical or adjacent loops at proximal positions suggesting feasible disulfide bonding, which can stabilize the domain.[19] Thus, both wild-type bias and tailored diversity were effective in producing an effective library. Additional engineering campaigns and sequence analysis will improve the statistical significance of these trends and guide further library improvement.

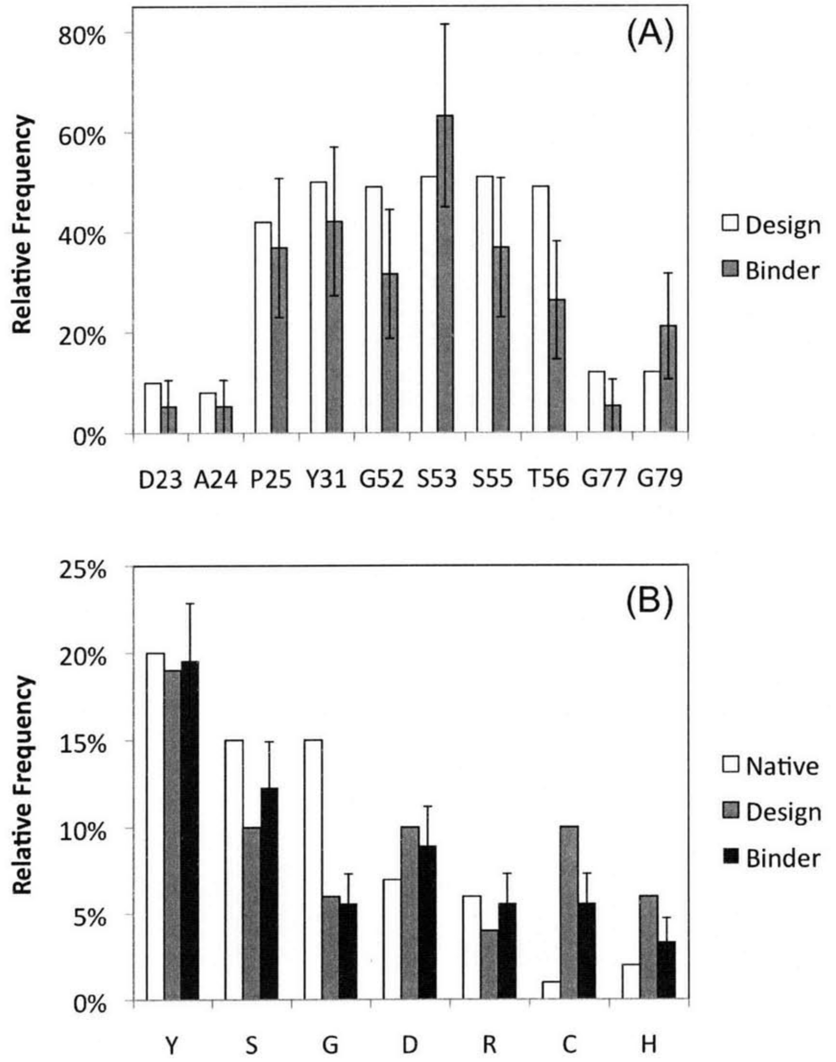


Figure 4.6. Binder sequence analysis. The nineteen binders from the G4 library were aligned and analyzed. (A) The wild-type frequency at each position with wild-type bias is indicated. (B) The amino acid frequency at positions without wild-type bias is indicated. *Native* indicates phylogenetic frequency. *Design* indicates the frequency in the G4 library design. *Binder* indicates the frequency in sequenced binders. The error bars represent a single standard deviation calculated as the square root of the counted amino acid occurrences divided by the total number of sequences.

Stability Analysis

The impact of wild-type bias and tailored diversity on domain stability was analyzed. The NNB and G4 libraries were each induced for yeast surface display at elevated temperature (37°). The G4 library exhibits 43±9% higher average display than the NNB library (Figure 4.7) indicating higher average stability. The libraries were then sorted by FACS to identify clones of low stability and high stability. About 50 clones were sequenced from each resultant population and the amino acid frequencies in low and high stability clones were compared (Table 4.4). The biased positions in the BC loop were not critical to stability in this analysis except position 29. As observed in binder sequence analysis, the small side chain alanine is preferred whereas the larger side chain leucine is destabilizing. Wild-type amino acids at the four biased positions in the DE loop are stabilizing, especially S53 and S55. While G77 is perhaps mildly stabilizing, G79 is present at substantially higher frequency in stable clones. The complete conservation of S85 in the G4 library is justified by the preferential occurrence of S85 in stable clones from the NNB library. At positions without wild-type bias, none of the preferred amino acids are substantially destabilizing thereby validating their inclusion at elevated levels.

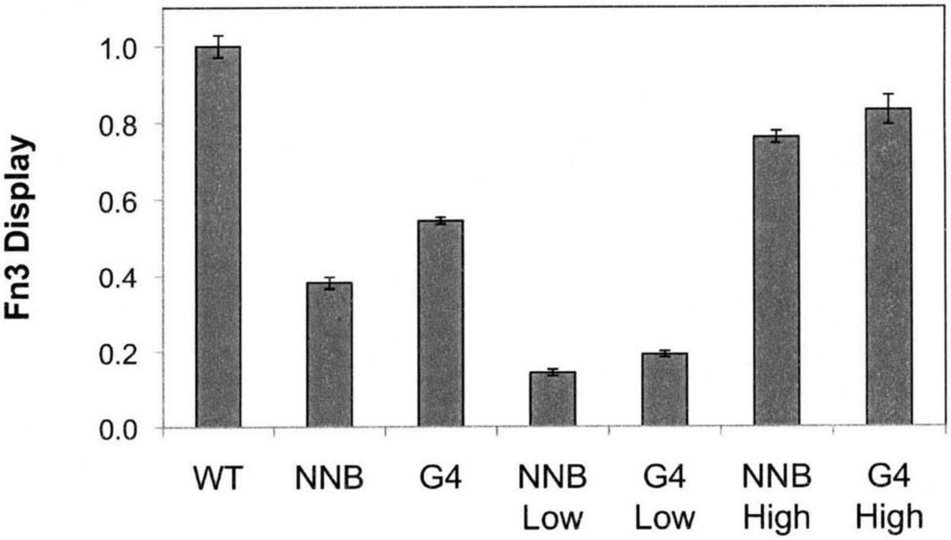


Figure 4.7. Library display. Yeast containing the indicated Fn3 populations were grown to logarithmic growth phase at 30°. Expression of Aga2p-Fn3 was induced at 37°. The mean Fn3 display level for each population was quantified by flow cytometry using mouse αc-myc antibody and αmouse antibody-AlexaFluor488 conjugate. *WT* is wild-type Fn3. *NNB* and *G4* are the naïve libraries. *Low* and *High* indicate the populations sorted for low and high display, respectively. Display levels are normalized to the wild-type value.

Table 4.4. Stability analysis. The NNB and G4 libraries were independently sorted for clones of low stability and high stability. Sequences of about 50 clones from each sorted population were analyzed. *AA* indicates the wild-type amino acid at positions with wild-type bias or amino acids of elevated frequency at positions without wild-type bias. *G4 Design* indicates the designed frequency of the indicated amino acid. *NNB* and *G4* indicate the difference in amino acid frequency between the high and low stability populations from the indicated library.

		High - Low (Stability)	
AA	G4 Design	NNB	G4
	Positions with Wild-type Bias		
D23	10% D	0%	+3%
A24	8% A	-4%	+1%
P25	42% P	0%	+4%
V29	A|L|S|V	+10|-19|+18|+5	+29|-11|-27|+12
Y31	S|Y	-	0%
G52	49% G	+7%	+10%
S53	51% S	+5%	+20%
S55	51% S	+18%	+44%
T56	49% T	+10%	+8%
G77	12% G	+1%	+5%
G79	12% G	+17%	+17%
S85	100% S	+18%	-
	Positions without Wild-type Bias		
Y	19%	-1%	-4%
S	10%	+5%	+2%
G	6%	0%	+1%
D	10%	+2%	+5%
R	4%	-1%	+2%
C	10%	1%	-2%
H	6%	0%	0%

Discussion

The current work demonstrates that tailored diversity is superior to nearly fully random (*e.g.*, NNB) or overly constrained (*e.g.*, YS) diversity. This is evidenced by the dominant selection of clones from the G4 library as well as the maintenance of the favored amino acids in binder sequences (Figure 4.6b). Tailored diversity improves the search of sequence space by increasing the frequency of functional binders. This results both through improving the likelihood of beneficial contacts, largely by elevation of tyrosine, and reducing detrimental constraints. The latter element is achieved through reduction of hydrophobic isoleucine, leucine, methionine, proline, threonine, and valine as well as the large, positively charged arginine and lysine, in deference to small, neutral serine. Yet a binary code of tyrosine and serine constrains sequence space such that it often lacks high affinity binders. Thus, through modest incorporation of other amino acids in the library and a broad, yet efficient mutagenesis approach, tailored diversity yields a vastly improved hybrid of the two extremes of NNB and YS.

The inclusion of wild-type bias is also an important element of the G4 library design. This bias increases the frequency of functional clones both by enabling diversity to be used at positions with more impact on binding and by reducing the number of non-functional clones that result from detrimental mutation of a structurally critical residue. Moreover, the improved stability of G4 clones (Figure 4.7) improves evolvability[3] allowing otherwise unstable sequence motifs to be explored. This improved stability is also beneficial in a variety of applications as outlined in the *Introduction*.

The methodology and techniques in the current work are directly applicable to any protein engineering effort. While the designed skewed nucleotide mixtures for particular sites are unique to Fn3, the antibody mimic mixture should be generally applicable to solvent-exposed loops in molecular recognition scaffolds. Moreover, the mixture design algorithm may be reapplied to any design distribution. The identification of positions most likely to benefit from wild-type bias can be readily applied to other scaffolds through high throughput stability analysis in the context of protein libraries, demonstrated here using yeast surface display. When available, sequence and structural data provide additional avenues of analysis. The relative efficacy of each of these approaches will be elucidated as continued analyses expand the sequence data set and evolved library designs are tested.

Though the thrust of this work entails study of sequence/structure/function relationships and library design, the panel of binders generated provides useful reagents for a variety of applications from tumor targeting (EGFR, human A33, and mouse A33) to biotechnology (HSA and mouse IgG) to immunology (FcγRIIa and FcγRIIIa). In addition, binders to tumor vasculature target CD276 were engineered solely from the G4 library (Appendix C).

Materials and Methods

Stability-Display Relationship

Yeast surface display plasmids were created for six Fn3 domains of previously published stabilities[12]: wild-type, 159, 159(wt DE), 159(Q8L), 159(A56E), and 159(Q8L,A56E). Genes were constructed by overlap extension PCR of eight oligonucleotides and transformed into EBY100 yeast as described.[18] Gene construction was verified by DNA sequencing. Clonal populations were grown at 30° in SD-CAA medium (0.07M sodium citrate pH 5.3, 6.7 g/L yeast nitrogen base, 5 g/L casamino acids, and 20 g/L glucose) and induced at 37° in SG-CAA (0.1M sodium phosphate, pH 6.0, 6.7 g/L yeast nitrogen base, 5 g/L casamino acids, 19 g/L galactose, and 1 g/L glucose). Yeast were labeled with mouse anti-c-myc antibody (clone 9E10) followed by phycoerythrin-conjugated goat anti-mouse antibody. Yeast were washed and phycoerythrin fluorescence was analyzed with an Epics XL flow cytometer (Beckman Coulter, Fullerton, CA).

Library Stability Analysis

A library was constructed in which positions 23-30 (DAPAVTVR), 52-55 (GSKST), and 77-86 (GRGDSPASSK) were diversified using NNB codons. The library was constructed by overlap extension PCR of eight oligonucleotides and transformed into EBY100 yeast. Fourteen similar libraries were constructed with identical design except a single codon of interest was maintained as wild-type within the otherwise diversified regions. Separate libraries were constructed for D23, A24, P25, A26, V27, T28, V29, G52, T56, G77, R78, G79, S84, and S85; in addition, a library was constructed that maintained D23, A24, P25, and V29. These libraries, as well as wild-type Fn3, were grown at 30° and induced at 37°; Fn3 expression was analyzed by flow cytometry as indicated above. The fractional improvement in display was calculated as the mean phycoerythrin fluorescence of the singly-conserved library minus that of the fully-diversified library and normalized to the fully-diversified fluorescence.

Solvent-Accessible Surface Area

The relative solvent accessible surface area of positions 22-32, 51-57, and 76-87 were calculated for wild-type Fn3 (solution structure 1TTG[14] and crystal structures 1FNA[15]) and an engineered binder (2OBG[16]). The area accessible to a 1.4Å sphere was determined for each side chain in each structure and compared to the accessible area in a G-X-G random coiled peptide using GetArea.[13]

Phylogenetic Sequence Alignment

The following fibronectin sequences were used: chimpanzee (XP_516072), cow (P07589), dog, (XP_536059), horse (XP_001489154), human (NP_997647), mouse (NP_034363), opossum (XP_001368449), platypus (XP_001509150), rat (NP_062016), and rhesus monkey (XP_001083548). The sequences were aligned using ClustalW.[22] The relative frequency of each amino acid was calculated at each position.

A similar analysis was conducted using engineered binder sequences. Engineered Fn3 domain sequences [7; 9; 10; 12; 16; 18; 19; 23-27] were aligned; identical loop sequences in related clones were only counted once to avoid bias. The amino acid frequency at each position was calculated and compared to the expected amino acid frequency as determined from a weighted average of theoretical library designs (*e.g.*, NNS, NNB, serine/tyrosine, etc.).

Library Construction

Degenerate oligonucleotides were designed to provide the desired amino acid distribution at each position. All three-site combinations of skewed nucleotide mixtures within 5%

increments were considered (*e.g.* 20% A, 5% C, 35% G, 40% T at the first position, 15% A, 45% C, 10% G, 30% T at the second position, and 35% A, 25% C, 30% G, 10% T at the third position). The amino acid probability distribution of each set of nucleotides mixtures was calculated from the genetic code. The sets were filtered to identify those with good tyrosine matching and reasonable matching of alanine, aspartic acid, glycine, arginine, and serine. Specifically, tyrosine was required to occur at 0.5-2x the intended frequency; alanine, aspartic acid, glycine, arginine, and serine were required to occur at 0.33-3x the intended frequency. The sets that fulfilled these criteria were then used to produce numerous *in silico* protein libraries based on their amino acid probability distribution. For each clone, the probability of occurrence from a library that precisely matched the desired distribution was calculated. The sum of probabilities for each sample library was used as a metric of library fitness. The skewed nucleotide designs were selected based on fitness and the ability to use identical mixtures at multiple sites (*e.g.*, 45% C, 10% G, 45% T at the wobble position of multiple codons). Nucleotide designs are included in Table 4.5.

Table 4.5. *Codon design.* The nucleotide mixture used in synthesis at each diversified position is indicated.

	Non-Conserved			A24			P25			V29			Y31		
	1st	2nd	3rd	1st	2nd	3rd	1st	2nd	3rd	1st	2nd	3rd	1st	2nd	3rd
A	0.15	0.45	0.00	0.15	0.45	0.00	0.05	0.20	0.00	0.00	0.00	0.00	0.00	0.50	0.00
C	0.15	0.15	0.45	0.15	0.30	0.45	0.65	0.65	0.45	0.00	0.50	0.00	0.00	0.50	0.00
G	0.25	0.25	0.10	0.25	0.15	0.10	0.05	0.05	0.10	0.50	0.00	1.00	0.00	0.00	0.00
T	0.45	0.15	0.45	0.45	0.10	0.45	0.25	0.10	0.45	0.50	0.50	0.00	1.00	0.00	1.00

	G52			S53 / S55			K54			T56			G77 / G79		
	1st	2nd	3rd	1st	2nd	3rd	1st	2nd	3rd	1st	2nd	3rd	1st	2nd	3rd
A	0.10	0.10	0.00	0.10	0.10	0.00	0.25	0.25	0.00	0.70	0.10	0.00	0.15	0.40	0.00
C	0.10	0.10	0.33	0.10	0.70	0.33	0.25	0.25	0.33	0.10	0.70	0.33	0.10	0.10	0.45
G	0.70	0.70	0.33	0.10	0.10	0.33	0.25	0.25	0.33	0.10	0.10	0.33	0.35	0.35	0.10
T	0.10	0.10	0.33	0.70	0.10	0.33	0.25	0.25	0.33	0.10	0.10	0.33	0.40	0.15	0.45

Degenerate oligonucleotides were synthesized with skewed nucleotides at diversified positions and nucleotides encoding wild-type Fn3 at fully-conserved positions. The library design, summarized in Table 4.1, includes four, three, and four loop lengths in the BC, DE, and FG loops. Separate oligonucleotides were synthesized to yield each length. Overlap extension PCR of eight oligonucleotides was performed to construct complete Fn3 genes. Separate reactions were conducted for each loop length to avoid bias towards shorter loops. The gene libraries were transformed into yeast by homologous recombination with linearized yeast surface display vector, which includes the Aga2p protein fusion, N-terminal HA epitope, and C-terminal c-myc epitope. The fraction of clones that produce full-length Fn3 was determined by flow cytometry as the fraction displaying the N-terminal HA tag that also contained the C-terminal c-myc epitope; these results were corroborated by sequence analysis.

Binder Selections

Human and mouse A33 extracellular domains were both produced with His_6 epitope tags in human embryonic kidney cells and purified by metal affinity chromatography. Protein was biotinylated either on free amines using the sulfo-NHS biotinylation kit or by site-specific sortase-based conjugation of GGGGG-biotin to an LPETG C-terminal epitope.[28] EGFR mutant 404SG[29] was produced in *Saccharomyces cerevisiae* yeast, purified by metal affinity chromatography and anti-EGFR antibody affinity chromatography, and biotinylated on free amines using the sulfo-NHS biotinylation kit. Biotinylated FcγRIIA and FcγRIIIA were a kind gift from Jeffrey Ravetch (Rockefeller University). Biotinylated mIgG was purchased from Rockland Immunochemicals. Human serum albumin (Sigma) was biotinylated using the sulfo-NHS biotinylation kit. The NNB, YS, and G4 libraries were pooled for direct competition.

The libraries were sorted for binding to the seven protein targets and affinity matured as described.[7] Yeast were grown and induced to display Fn3. Binders to streptavidin-coated magnetic Dynabeads were removed.[21] Biotinylated protein was loaded on streptavidin-coated magnetic Dynabeads and incubated with the remaining yeast. The beads were washed with PBSA and the beads with attached cells were grown for further selection. After two magnetic bead sorts, full-length Fn3 clones were selected by fluorescence-activated cell sorting using the C-terminal c-myc epitope for identification of full-length clones. Plasmid DNA was zymoprepped from the cells and mutagenized by error-prone PCR of the entire Fn3 gene or the BC, DE, and FG loops. Mutants were transformed into yeast by electroporation with homologous recombination and requisite shuffling of the loop mutants. The lead clones and their mutants were pooled for further cycles of selection and mutagenesis. Once significant binder enrichment was observed during magnetic bead sorts, fluorescence activated cell sorting was used. Yeast displaying Fn3 were incubated with biotinylated target protein and anti-c-myc antibody (clone 9E10 or chicken anti-c-myc, Invitrogen). Cells were washed and incubated with AlexaFluor488-, phycoerythrin-, or AlexaFluor647-conjugated streptavidin and fluorophore-conjugated anti-mouse or anti-chicken antibody. Cells were washed and cells with the highest target to c-myc labeling ratio were selected on a FACS Aria or MoFlo flow cytometer. Plasmids from binding populations were zymoprepped and transformed into *E. coli*; transformants were grown, miniprepped, and sequenced.

Library Source Determination

For each clone, the probabilities that it originated from the NNB, YS, or G4 library were calculated using the designed nucleotide distributions at each position as well as the probability of mutation by error-prone PCR.

Library Stability Analysis

The NNB and G4 libraries were independently grown at 30° and induced at 37°. Yeast were labeled with mouse anti-HA antibody (clone 16B12, Covance) and chicken anti-c-myc antibody to label the N- and C-terminal epitopes. Cells were washed, incubated with phycoerythrin-conjugated goat anti-mouse antibody and AlexaFluor488-conjugated goat anti-chicken antibody, and sorted by flow cytometry. Only cells were comparable signals for each epitope were considered to avoid selecting epitope mutants. The lowest and highest displaying cells were collected and grown for an additional induction and selection. Plasmids were isolated and transformed into *E. coli*. About 50 clones from each resultant population (both low and high stability for both NNB and G4) were miniprepped and sequenced. Sequences were aligned and the amino acid frequencies at each position were determined.

References

1. Binz, H., Amstutz, P. & Plückthun, A. (2005). Engineering novel binding proteins from nonimmunoglobulin domains. *Nat Biotechnol* **23**, 1257-1268.
2. Sidhu, S. & Fellouse, F. (2006). Synthetic therapeutic antibodies. *Nat Chem Biol* **2**, 682-688.
3. Bloom, J. D., Labthavikul, S. T., Otey, C. R. & Arnold, F. H. (2006). Protein stability promotes evolvability. *Proc Natl Acad Sci USA* **103**, 5869-74.
4. Fellouse, F., Wiesmann, C. & Sidhu, S. (2004). Synthetic antibodies from a four-amino-acid code: a dominant role for tyrosine in antigen recognition. *Proc Natl Acad Sci USA* **101**, 12467-72.
5. Fellouse, F., Li, B., Compaan, D. M., Peden, A. A., Hymowitz, S. G. & Sidhu, S. (2005). Molecular recognition by a binary code. *Journal of Molecular Biology* **348**, 1153-62.
6. Fellouse, F., Barthelemy, P. A., Kelley, R. F. & Sidhu, S. (2006). Tyrosine plays a dominant functional role in the paratope of a synthetic antibody derived from a four amino acid code. *Journal of Molecular Biology* **357**, 100-14.
7. Hackel, B. J. & Wittrup, K. D. *submitted.*
8. Fellouse, F., Esaki, K., Birtalan, S., Raptis, D., Cancasci, V. J., Koide, A., Jhurani, P., Vasser, M., Wiesmann, C., Kossiakoff, A. A., Koide, S. & Sidhu, S. (2007). High-throughput generation of synthetic antibodies from highly functional minimalist phage-displayed libraries. *Journal of Molecular Biology* **373**, 924-40.
9. Gilbreth, R. N., Esaki, K., Koide, A., Sidhu, S. & Koide, S. (2008). A dominant conformational role for amino acid diversity in minimalist protein-protein interfaces. *Journal of Molecular Biology* **381**, 407-18.
10. Huang, J., Koide, A., Makabe, K. & Koide, S. (2008). Design of protein function leaps by directed domain interface evolution. *Proc Natl Acad Sci USA.*
11. Shusta, E. V., Kieke, M. C., Parke, E., Kranz, D. M. & Wittrup, K. D. (1999). Yeast polypeptide fusion surface display levels predict thermal stability and soluble secretion efficiency. *Journal of Molecular Biology* **292**, 949-56.
12. Parker, M., Chen, Y., Danehy, F., Dufu, K., Ekstrom, J., Getmanova, E., Gokemeijer, J., Xu, L. & Lipovsek, D. (2005). Antibody mimics based on human fibronectin type three domain engineered for thermostability and high-affinity binding to vascular endothelial growth factor receptor two. *Protein Engineering Design and Selection* **18**, 435-444.
13. Fraczkiewicz, R. & Braun, W. (1998). Exact and efficient analytical calculation of the accessible surface areas and their gradients for .... *Journal of computational chemistry.*
14. Main, A. L., Harvey, T. S., Baron, M., Boyd, J. & Campbell, I. D. (1992). The three-dimensional structure of the tenth type III module of fibronectin: an insight into RGD-mediated interactions. *Cell* **71**, 671-8.
15. Dickinson, C. D., Veerapandian, B., Dai, X. P., Hamlin, R. C., Xuong, N. H., Ruoslahti, E. & Ely, K. R. (1994). Crystal structure of the tenth type III cell adhesion module of human fibronectin. *Journal of Molecular Biology* **236**, 1079-92.
16. Koide, A., Gilbreth, R. N., Esaki, K., Tereshko, V. & Koide, S. (2007). High-affinity single-domain binding proteins with a binary-code interface. *Proc Natl Acad Sci USA* **104**, 6632-7.
17. Birtalan, S., Zhang, Y., Fellouse, F., Shao, L., Schaefer, G. & Sidhu, S. (2008). The intrinsic contributions of tyrosine, serine, glycine and arginine to the affinity and specificity of antibodies. *Journal of Molecular Biology* **377**, 1518-28.
18. Hackel, B., Kapila, A. & Wittrup, K. (2008). Picomolar affinity fibronectin domains engineered utilizing loop length diversity, recursive mutagenesis, and loop shuffling. *Journal of Molecular Biology* **381**, 1238-52.
19. Lipovsek, D., Lippow, S., Hackel, B., Gregson, M. W., Cheng, P., Kapila, A. & Wittrup, K. (2007). Evolution of an interloop disulfide bond in high-affinity antibody mimics based on fibronectin type III domain and selected by yeast surface display: molecular convergence with single-domain camelid and shark antibodies. *Journal of Molecular Biology* **368**, 1024-41.
20. Zemlin, M., Klinger, M., Link, J., Zemlin, C., Bauer, K., Engler, J. A., Schroeder, H. W. & Kirkham, P. M. (2003). Expressed murine and human CDR-H3 intervals of equal length exhibit distinct repertoires that differ in their amino acid composition and predicted range of structures. *Journal of Molecular Biology* **334**, 733-49.

21. Ackerman, M., Levary, D., Tobon, G., Hackel, B., Orcutt, K. D. & Wittrup, K. (2009). Highly avid magnetic bead capture: an efficient selection method for de novo protein engineering utilizing yeast surface display. *Biotechnol Prog* **25**, 774-83.
22. Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., F., V., M., W. I., A., W., Lopez, R., J.D., T., T.J., G. & D.G., H. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics*.
23. Koide, A., Bailey, C. W., Huang, X. & Koide, S. (1998). The fibronectin type III domain as a scaffold for novel binding proteins. *Journal of Molecular Biology* **284**, 1141-51.
24. Koide, A., Abbatiello, S., Rothgery, L. & Koide, S. (2002). Probing protein conformational changes in living cells by using designer binding proteins: application to the estrogen receptor. *Proc Natl Acad Sci USA* **99**, 1253-8.
25. Xu, L., Aha, P., Gu, K., Kuimelis, R. G., Kurz, M., Lam, T., Lim, A. C., Liu, H., Lohse, P. A., Sun, L., Weng, S., Wagner, R. W. & Lipovsek, D. (2002). Directed evolution of high-affinity antibody mimics using mRNA display. *Chemistry & Biology* **9**, 933-42.
26. Karatan, E., Merguerian, M., Han, Z., Scholle, M. D., Koide, S. & Kay, B. K. (2004). Molecular recognition properties of FN3 monobodies that bind the Src SH3 domain. *Chemistry & Biology* **11**, 835-44.
27. Olson, C. A., Liao, H. I., Sun, R. & Roberts, R. W. (2008). mRNA display selection of a high-affinity, modification-specific phospho-IkappaBalpha-binding fibronectin. *ACS Chem Biol* **3**, 480-5.
28. Parthasarathy, R., Subramanian, S. & Boder, E. T. (2007). Sortase A as a novel molecular "stapler" for sequence-specific protein conjugation. *Bioconjug Chem* **18**, 469-76.
29. Kim, Y., Bhandari, R., Cochran, J., Kuriyan, J. & Wittrup, K. (2006). Directed evolution of the epidermal growth factor receptor extracellular domain for expression in yeast. *Proteins* **62**, 1026-1035.

5. EPIDERMAL GROWTH FACTOR RECEPTOR DOWNREGULATION WITH BIVALENT FIBRONECTIN CONSTRUCTS

Introduction

EGFR is a validated cancer target manifested by dysregulation[1], overexpression,[2] autocrine signaling,[3] and mutation.[4-6] Yet the FDA-approved ligand-blocking antibodies cetuximab and panitumumab have only modest efficacy.[7-9] Deficient performance could result from inability to compete with autocrine ligand, insufficient downregulation of receptor, or inactivity against mutants such as constitutively active EGFRvIII.

An alternative mode of therapy is substantial receptor downregulation to reduce or eliminate the detrimental effects of receptor activation on tumor formation, proliferation, and migration. A previously demonstrated means of receptor downregulation is administration of non-competitive pairs of antibodies. Antibodies 528 and 806 downregulate EGFR and synergistically inhibit tumor xenografts.[10] Non-competitive antibody pairs 111 + 565 and 143 + 565 downregulate EGFR whereas the competitors 111 + 143 do not.[11] Also, non-competitive anti-HER2 antibodies downregulate HER2 and inhibit tumor growth.[11; 12] However, these approaches require dosing two molecules, which complicates regulatory and clinical procedures; moreover, decoupled pharmacokinetics could reduce synergy. A bispecific molecule could potentially alleviate these problems though the efficacy is uncertain given the lack of mechanistic detail in the published literature. Fn3 domains provide a good system for bispecific constructs because their single-domain architecture enables simple head-to-tail fusion, which is the natural state of Fn3 domains within complete fibronectin protein.

In the current work, we engineer a panel of small, single-domain EGFR binders to multiple identified receptor epitopes. Homo- and hetero-bivalent combinations of these binders, expressed as protein fusions, are tested for the ability to downregulate receptor in a variety of cell lines. Several molecules effectively reduce EGFR levels up to 80%. The impact of epitopes, receptor density, bivalent format, and avidity are investigated. Phosphorylation, both of receptor and downstream molecules, is examined. Inhibition of proliferation and migration through downregulation is demonstrated.

Results

Binder Engineering

Multiple high affinity binders to distinct epitopes of EGFR ectodomain were desired. The NNB, YS, and G4 libraries were pooled and sorted for binding to biotinylated EGFR ectodomain mutant 404SG.[13] Two clones dominated the selection. Competition against existing anti-EGFR antibodies revealed that clone E4.2.2 is competitive with ICR10, a domain I binder, and clone E4.2.1 is competitive with 528, a domain III binder. To identify additional binders, intermediate populations were sorted for binding to EGFR ectodomain in the presence of ICR10 or 528. Five unique clones that bound ICR10-blocked EGFR were identified: EI4.4.2, EI3.4.3, EI3.4.2, EI2.4.6, and EI1.4.1. Also, two additional rounds of sorting with unblocked EGFR yielded an improved mutant of E4.2.2 named E6.2.6 and one additional clone, E6.2.10 (Table 5.1). In addition to binding soluble EGFR ectodomain produced in yeast, these eight clones all bind EGFR-expressing human epidermoid carcinoma A431 cells (data not shown). The affinity of each clone was determined by titration of biotinylated Fn3 binding to A431 (on ice to prevent internalization); affinities ranged from 250 pM to 30 nM (Table 5.1, Figure 5.1).

Table 5.1. *EGFR binders. K_d* indicates equilibrium dissociation constant for binding to A431 cells on ice or yeast at 22°. *nb* indicates no detectable binding. – indicates data not collected.

		Sequence				K_d [nM]		
Name	Alias	BC	DE	FG	fw	A431, pH 7.4	A431, pH 5	Yeast, pH 7.4
Wild-type	WT	DAPAVTVRY	GSKST	GRGDSPASSK	-	nb	nb	nb
E6.2.6	A	FDYAVTY	GWIST	DNSHWPFRST	I90T	0.26 ± 0.13	0.26 ± 0.14	1.2 ± 0.4
E4.2.1	B	YGFSLASS	RSPWF	SNDFSNRYSG	-	30 ± 3	2.5 ± 0.7	0.25 ± 0.07
EI3.4.2	-	YGSSYASY	RSPWF	PSGISA	T58I	9.5 ± 3.5	0.80 ± 0.28	1.0 ± 0.2
E6.2.10	-	YLRDPRYVDY	WYLPE	YDGYRESTPL	-	0.96 ± 0.11	0.88 ± 0.64	-
EI4.4.2	C	YFRDPRYVDY	WYLPE	GDDQNAGL	V45A	1.4 ± 0.2	0.64 ± 0.32	4 ± 4
EI3.4.3	D	LHHRSDVRS	GSRSL	WGSYCCSN	E47K	0.25 ± 0.05	0.081 ± 0.044	2.5 ± 0.1
EI2.4.6	E	YHPFYYVAHS	RSPWF	DSNGSH	-	2.9 ± 0.3	nb	4 ± 4
EI1.4.1	-	YGPFYYVAHS	RSPWF	SKCYDGSV	-	0.85 ± 0.50	~0.15	0.06 ± 0.04

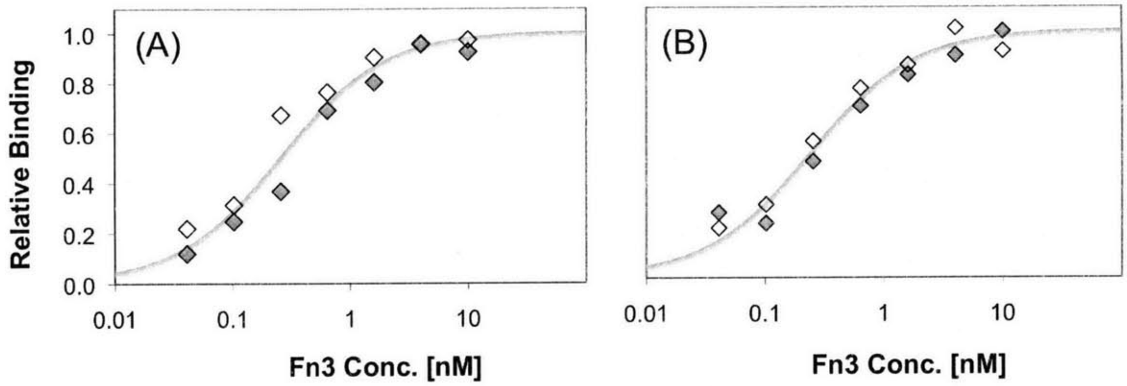


Figure 5.1. *Affinity titrations.* A431 cells were incubated with the indicated concentration of biotinylated E6.2.6 (A) or EI3.4.3 (B), washed, labeled with streptavidin-R-phycoerythrin, and analyzed by flow cytometry. Relative binding indicates the mean fluorescence normalized between minimum and maximum signal.

Competition and Epitope Mapping

Clones A-E , EI3.4.2, and EI1.4.1 bind conformationally-sensitive epitopes as evidenced by their inability to bind EGFR ectodomain after thermal denaturation of receptor on the yeast surface (Figure 5.2). Binders were tested for the ability to compete with other clones as well as antibodies 225, 528, and ICR10 (Figure 5.3). Clone A is competitive solely with ICR10, a known domain I binder.[14] This result was corroborated by the ability of clone A to bind the EGFR ectodomain fragment comprising amino acids 1-176 displayed on the yeast surface. Clone D is not competitive with the other Fn3s or antibodies tested. It is able to bind ectodomain fragments 294-543 and 302-503, thereby localizing the binding to domain III and the beginning of domain IV. Clones B, C, E, EI3.4.2, and EI1.4.1 compete with each other as well as antibodies 225 and 528, EGF-competitive domain III binders (except for three untested combinations; see Figure 5.3). Clones A-E, as well as E6.2.10, compete with EGF for binding to A431 cells.

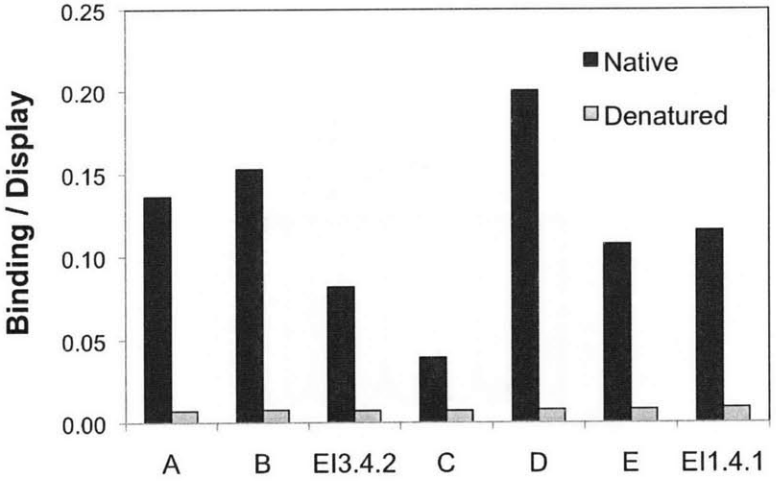
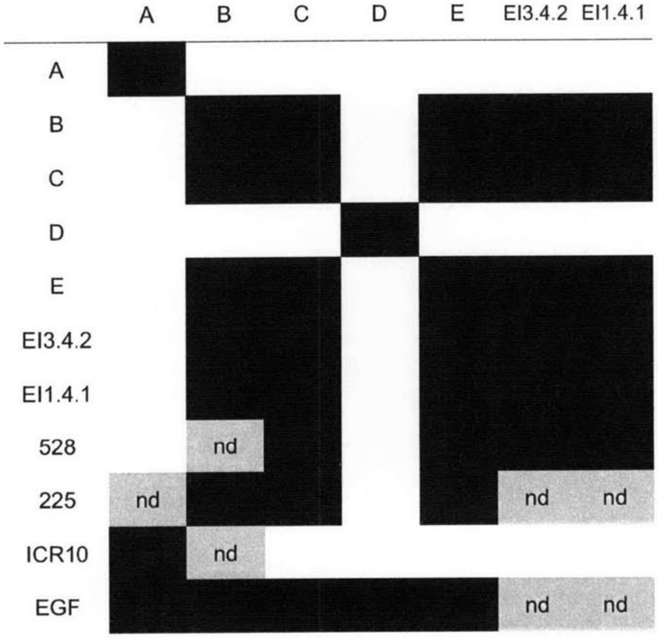


Figure 5.2. *Conformational sensitivity.* EGFR ectodomain mutant 404SG was displayed on the yeast surface. Cells were incubated at 80° for 30 min. to denature EGFR. Cells were labeled with biotinylated Fn3 and mouse anti-c-myc antibody followed by streptavidin-R-phycoerythrin and AlexaFluor488-conjugated anti-mouse antibody. Fluorescence was quantified by flow cytometry.



	A	B	C	D	E	EI3.4.2	EI1.4.1
A	■						
B		■	■		■	■	■
C		■	■		■	■	■
D				■			
E		■	■		■	■	■
EI3.4.2		■	■		■	■	■
EI1.4.1		■	■		■	■	■
528		nd	■		■	■	■
225	nd	■	■		■	nd	nd
ICR10	■	nd					
EGF	■	■	■	■	■	nd	nd

Figure 5.3. *Binding Competition.* A431 cells (for 225 and EGF competition) were incubated on ice with the indicated Fn3 clone or PBSA control. AlexaFluor488-conjugates of 225 or EGF were added and cells were analyzed by flow cytometry. For all other competitions, yeast displaying EGFR ectodomain were incubated Fn3 clone, 528, or ICR10 followed by biotinylated Fn3, which was detected by streptavidin-R-phycoerythrin and flow cytometry. Black indicates competition. White indicates no competition. *nd* indicates samples that were not determined.

Higher resolution epitope mapping was performed by high throughput identification of EGFR mutations that maintain foldedness but have reduced affinity for the clone of interest.[15] In agreement with competition and fragment labeling, clone A binds to domain I as evidenced by its reduced binding to mutants L14H, Q16R, Y45F, and H69(QRY) (Figure 5.4A). The specific location in domain I provides an explanation for EGF competition as the four sites identified for clone A binding are all within 4Å of EGF in the EGF/EGFR crystal structure (Figure 5.4B). Clones B, C, E, and E6.2.10 all bind domain III on the portion closer to domain II, which is consistent with complementary Fn3 competition as well as EGF competition. Antibody 225 competition is reasonable for clones B, C, and E given their proximity to the cetuximab (a 225 chimera) interface (Figure 5.4C). The lack of E6.2.10 competition with 225 binding is also acceptable given their disparate, though proximal, epitopes. Clone D binds near the interface of domains III and IV, which is consistent with its fragment labeling and lack of competition against 225 and clones B, C, and E. The ability of clone D to compete with EGF cannot readily be explained by direct steric inhibition given their distal binding epitopes. However, a reasonable hypothesis is that clone D binding inhibits receptor untethering that supports high affinity ligand binding. Though domains III and IV do not grossly change during untethering,[16] subtle rearrangements at the domain III / domain IV interface exist; for example, amino acids 430 and 506, which are the sites identified in clone D epitope mapping, move from 19.7Å apart in the tethered structure to 16.7Å in the dimer.

Thus, at least three classes of binders have been engineered: clone A binds to domain I; clones B, C, and E bind domain III and are competitive with each other and antibodies 225 and 528 (as well as EI1.4.1 and EI3.4.2); clone D binds to the C-terminal portion of domain III and the N-terminal portion of domain IV and does not compete with antibodies 225 and 528 nor clones B, C, E, EI1.4.1, and EI3.4.2.

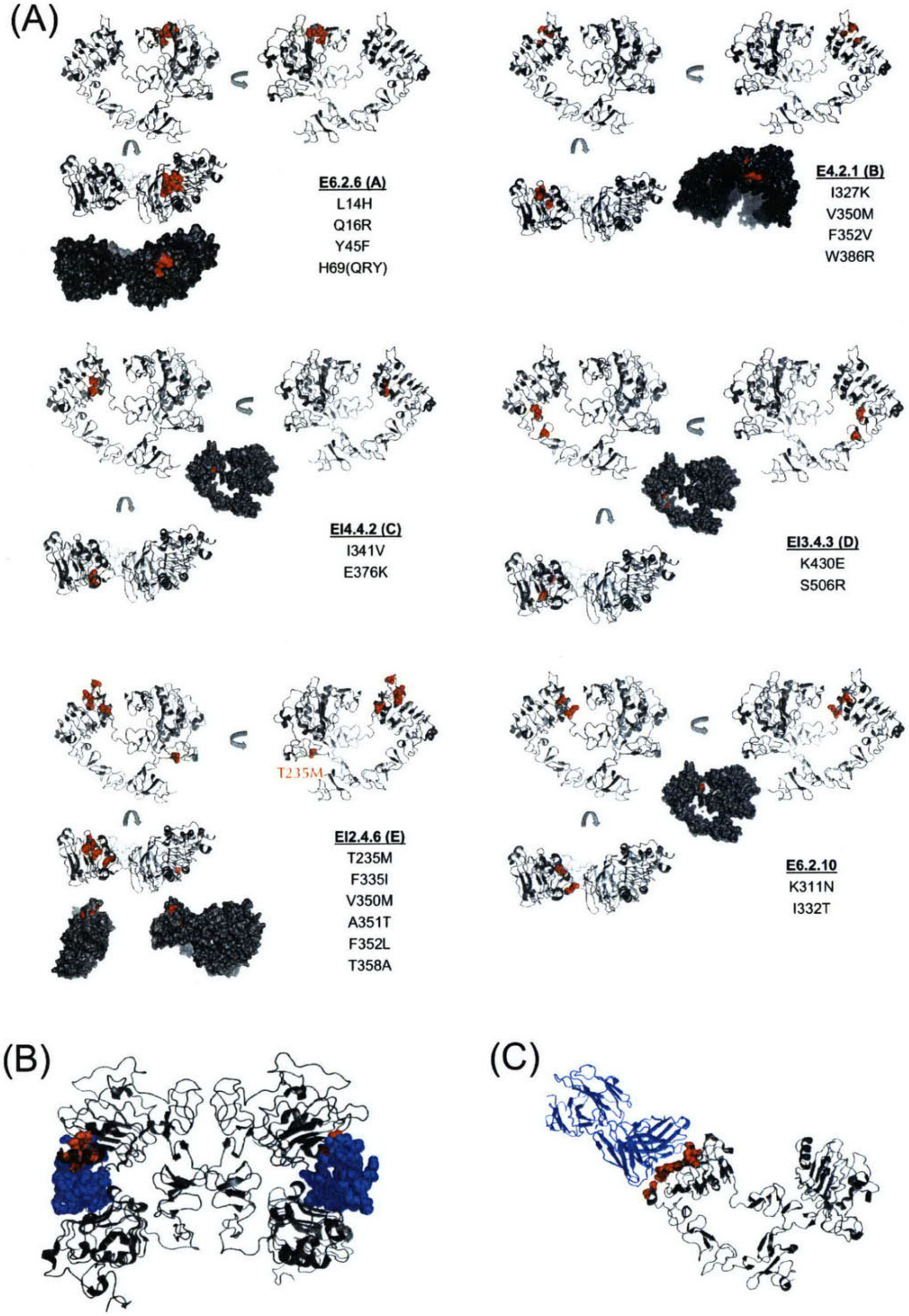

(A)
E6.2.6 (A)
L14H
Q16R
Y45F
H69(QRY)
E4.2.1 (B)
I327K
V350M
F352V
W386R
EI4.4.2 (C)
I341V
E376K
EI3.4.3 (D)
K430E
S506R
T235M
EI2.4.6 (E)
T235M
F335I
V350M
A351T
F352L
T358A
E6.2.10
K311N
I332T
(B)
(C)

Figure 5.4. *Fine epitope mapping.* (A) A library of EGFR ectodomain mutants was sorted for clones that maintained binding to a conformational binder but had reduced binding to the indicated Fn3 domain. All single amino acid mutants, excluding proline and glycine mutants, are listed and presented as red spheres in the ectodomain crystal structure (1NQL). (B) EGF bound to EGFR dimer (1IVO[17]). EGF is shown in blue spheres. Clone A epitope shown in red spheres. (C) Cetuximab (a chimera of 225) bound to EGFR in the tethered conformation (1YY9[18]). Residues within 4Å of antibody are shown in red spheres.

Downregulation by Heterobivalent Constructs

Given the previously reported success of particular pairs of non-competitive homobivalent antibodies to downregulate EGFR, we sought to investigate the ability to achieve similar downregulation via a single heterobivalent agent. Fn3 clones were linked as head-to-tail protein fusions with the native seven amino acid EIDKSPQ as well as a flexible GSGGGSGGGKGGGGT linker (Figure 5.5A). Thirty constructs comprising all possible bivalent combinations, in both orientations, as well as monomer for five clones (identified as A-E under *Alias* in Table 5.1; bivalents are named N-C where *N* and *C* represent the N-terminal and C-terminal Fn3 clones) were tested. Three different EGFR-expressing human cell lines were tested: A431 epidermoid carcinoma, HeLa cervical carcinoma, and HT29 colorectal carcinoma. Cells were cultured, serum starved, and incubated with 20 nM Fn3 or Fn3-Fn3 for 6-8h. Cells were detached, bound agent was acid stripped, and surface EGFR was quantified by flow cytometry. Although many constructs did not modify surface EGFR levels relative to PBSA control, bivalents D-B, D-C, D-D, D-E, A-D, B-D, C-D, and E-D downregulate, yielding up to 80% reduction in surface EGFR; D-B, D-C, and D-E have the greatest effect (Figure 5.5B,C). Thus, particular combinations of non-competitive clones in a heterobivalent construct are needed to downregulate though the D-D homobivalent does moderately reduce receptor levels. Moreover, particular orders of combinations are needed; for example A-D downregulates whereas D-A does not.

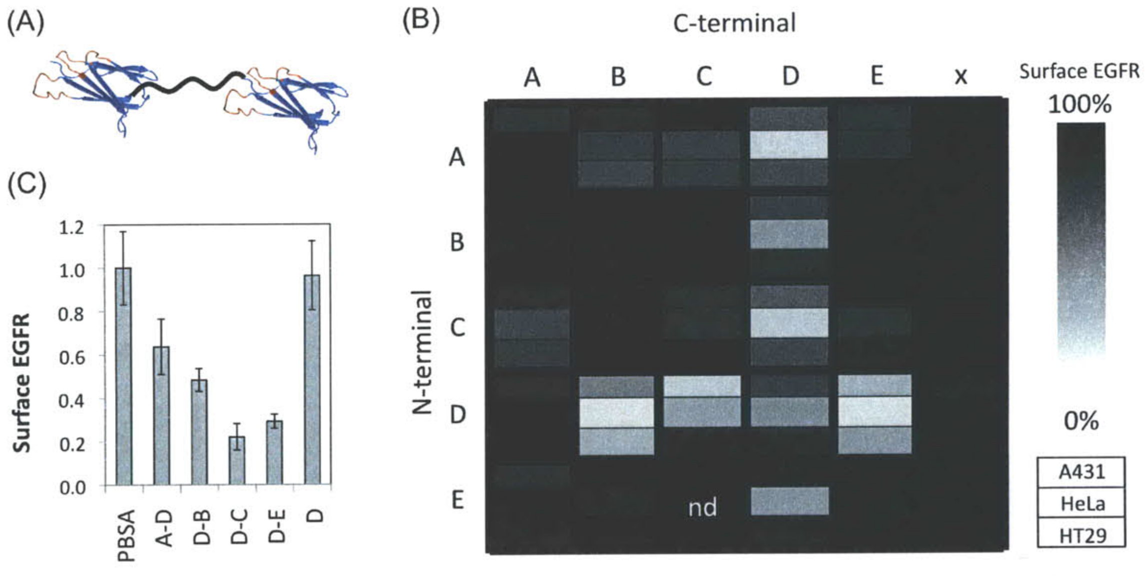


Figure 5.5. *EGFR downregulation.* (A) Schematic of Fn3-Fn3 heterobivalent with the wild-type Fn3 structure from PDB ID 1TTG and the flexible linker drawn approximately to scale in cartoon form. (B) A431, HeLa, and HT29 cells were cultured in 96-well plates, serum starved, and treated with 20 nM of the indicated Fn3 or Fn3-Fn3 construct for 6-8h. Surface EGFR was quantified by flow cytometry and is presented on a color scale relative to PBSA-treated control with black indicating no downregulation and white indicating complete downregulation. Mean of triplicate samples is used for quantification. (C) Data from (B) for select constructs with A431 cells. Error bars indicate standard deviation of triplicate samples.

Multiple elements of downregulation were investigated. To further expand the generality of downregulation efficacy as well as to examine the impact of receptor density, three heterobivalents were tested on additional cell lines: U87 glioblastoma, hMEC (human mammary epithelial cells), and Chinese hamster ovary (CHO) cells transfected with EGFR-green fluorescent protein fusion. Downregulation was observed in all six cell lines for D-B, D-C, and D-E (Figure 5.6). Interestingly, downregulation was reduced for D-C and D-E in the low-expressing cells HT29 and U87. Conversely, EGF downregulates receptor most robustly in these low-expressing lines while exhibiting muted receptor reduction in the high-expressing CHO and A431 cells.

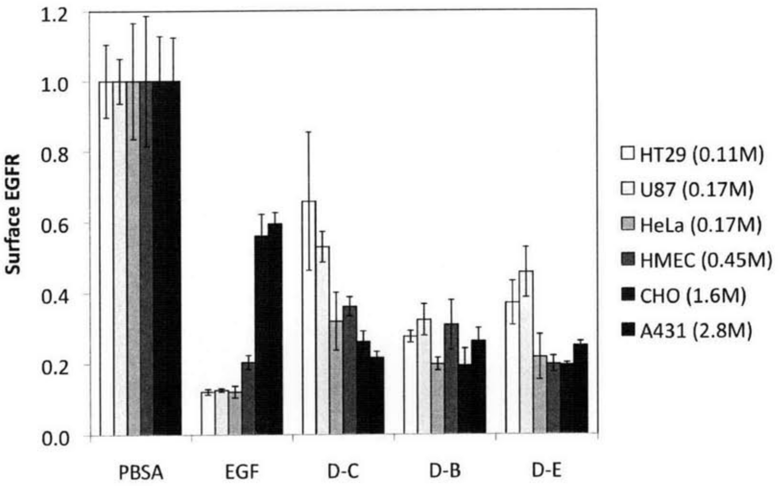


Figure 5.6. *Downregulation in various cell lines.* Cells were cultured in 96-well plates, serum starved, and treated with 20 nM agent for 8h. Surface EGFR was quantified by flow cytometry and normalized to PBSA-treated control. Values and error bars indicate the mean and standard deviation of triplicate samples. Parenthetical notation in legend indicates the number of EGFR per cell in million (M).

Downregulation kinetics were analyzed for the most robust heterobivalents. D-B and D-C downregulate EGFR in A431 cells with half-times of 1.1 and 1.4h, respectively (Figure 5.7). Downregulation in HeLa cells is slightly faster at 0.44, 0.59, and 1.3h for D-B, D-C, and D-E.

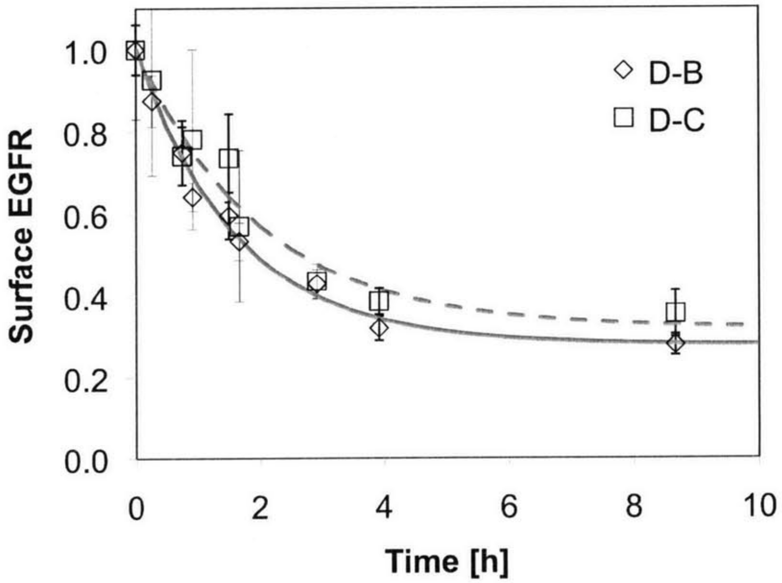


Figure 5.7. *Downregulation kinetics.* Cells were cultured in 96-well plates, serum starved, and treated with 20 nM D-B (triangles) or D-C (squares) for the indicated time. Surface EGFR was quantified by flow cytometry and normalized to PBSA-treated control. Values and error bars indicate the mean and standard deviation of triplicate samples.

Heterobivalent D-C and D-E constructs were created with three different lengths of the linker between the Fn3 domains; in addition to the native EIDKPSQ glycine-rich linkers of four, 15, or 27 amino acids were included. These constructs were tested for downregulation of EGFR in HT29, U87, HeLa, hMEC, CHO, and A431 cells. Although results vary by cell line and heterobivalent, the long linker is always the least effective and the shortest linker is often the most effective (Figure 5.8).

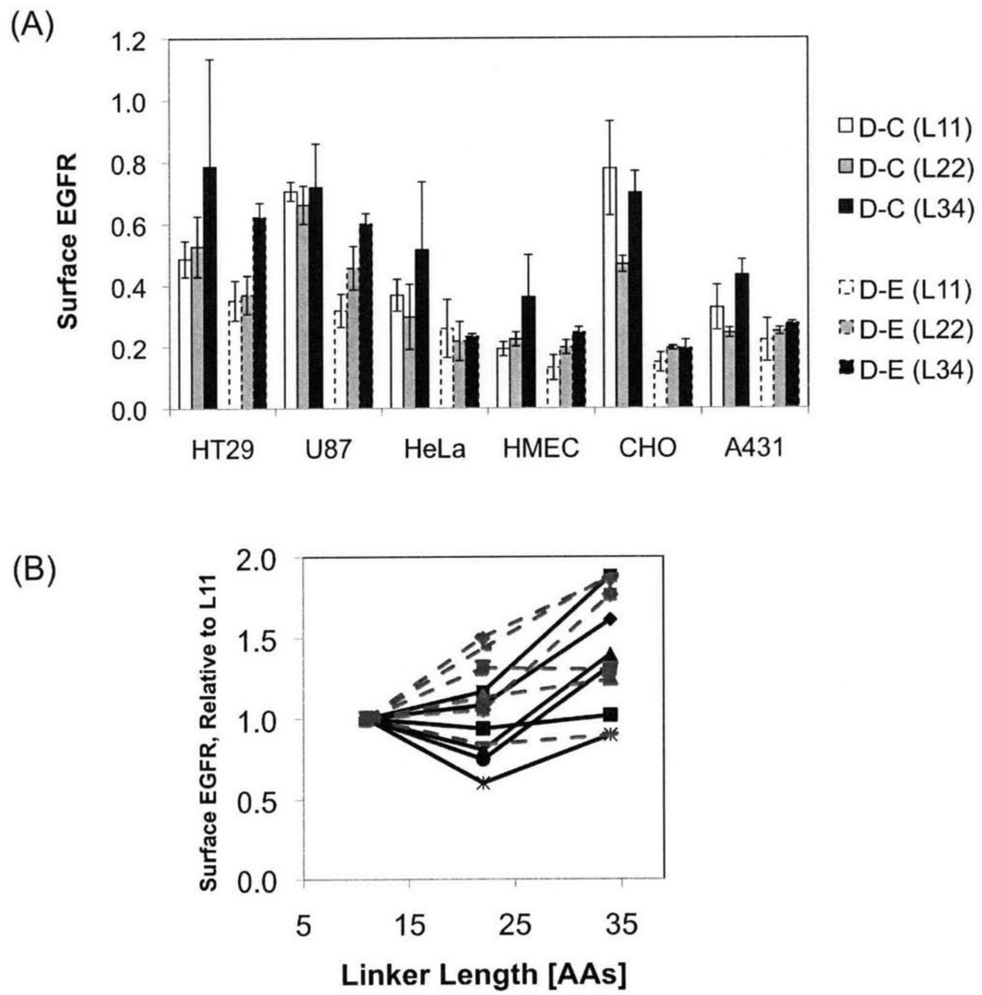


Figure 5.8. *Linker length effect.* (A) Cells were cultured in 96-well plates, serum starved, and treated with 20 nM D-C or D-E with the indicated linker length for 8h. Surface EGFR was quantified by flow cytometry and normalized to PBSA-treated control. Values and error bars indicate the mean and standard deviation of triplicate samples. (B) The data from (A) is summarized to compare linker lengths. Surface EGFR values are normalized for each combination (D-C or D-E) and cell type.

An alternative format of bispecific was tested in which monovalent Fn3 domains were biotinylated and combinations of clones were immobilized on AlexaFluor488-conjugated streptavidin. In all bispecific and trispecific combinations of A, C, D, E, EI1.4.1, and EI3.4.2, no downregulation is observed in HT29 or U87 cells transfected to overexpress EGFR (data not shown). Yet most combinations yield a substantial accumulation of internalized AlexaFluor488 signal suggestive of complex internalization without downregulation. Thus, bispecific format appears critical for efficacy. Of note, internalized AlexaFluor488 signal at 37° correlates with surface labeling at 4° (which restricts internalization) suggestive of passive internalization for all combinations.

Phosphorylation

To investigate the mechanisms of downregulation, an EGFR expression vector was transfected into human embryonic kidney (HEK) cells, which express low levels of native EGFR. Though EGF robustly downregulates native HEK EGFR, transfected cells with approximately 50-fold more EGFR are not effectively downregulated. Conversely, D-B and D-C heterobivalents are able to downregulate transfected EGFR (Figure 5.9A). The activity of the transfected EGFR is validated by a strong correlation between the fraction of cells transfected and the downregulation of native EGF (Figure 5.9B); thus, the presence of overexpressing transfected cells reduces the EGF-based downregulation of non-transfected cells possibly through ligand depletion or competition. These results indicate a divergence between the mechanisms of downregulation by EGF and Fn3-Fn3 heterobivalents.

To further explore the mechanism, eight EGFR mutants with point mutations in their intracellular domains were tested for their ability to be downregulated. All eight mutants (T654A, T669A, K721R, Y845F, S1046A/S1047A, Y1068F, Y1148F, Y1173F) exhibit downregulation on par with wild-type EGFR in the presence of D-B and D-C (Figure 5.9C).

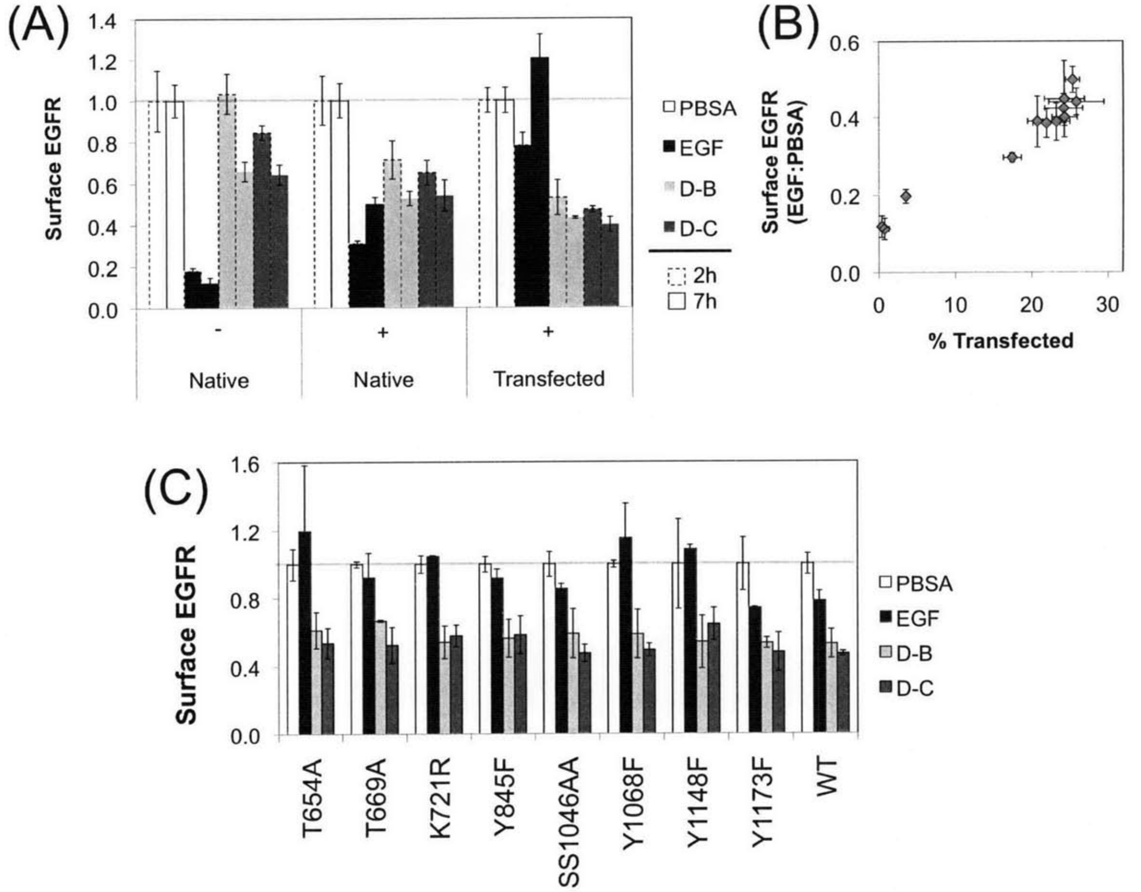


Figure 5.9. *Downregulation of HEK transfectants.* HEK cells were transfected with an EGFR expression vector, grown, and treated with 20 nM agent for 2 or 7h. Surface EGFR was quantified by flow cytometry and normalized to PBSA-treated control. (A) Wild-type EGFR transfection. + and – indicate presence or absence of transfection vector in the sample well. *Native* indicates analysis of cells that were not successfully transfected. *Transfected* indicates analysis of cells that were successfully transfected. (B) The surface EGFR level in EGF-treated samples relative to PBSA-treated samples is plotted for samples of various transfection efficiencies for the 7h treatment. (C) EGFR mutants were transfected and cells were treated for 2h prior to analysis. Values and error bars represent mean and standard deviation of at least quadruplicate samples.

The impact of heterobivalents on EGFR phosphorylation was analyzed at eight sites: T654, T669, Y845, S1046, Y1068, Y1086, Y1148, and Y1173. Heterobivalent D-C, PBSA, or EGF was added to A431 cells for 5, 15, 60, or 240 min. and receptor phosphorylation was quantified by in-cell Western blot. Receptor agonism by D-C is consistently lower than that by EGF with the lone exception of T669 at early times (Figure 5.10). In fact, receptor agonism is often non-distinct from background.

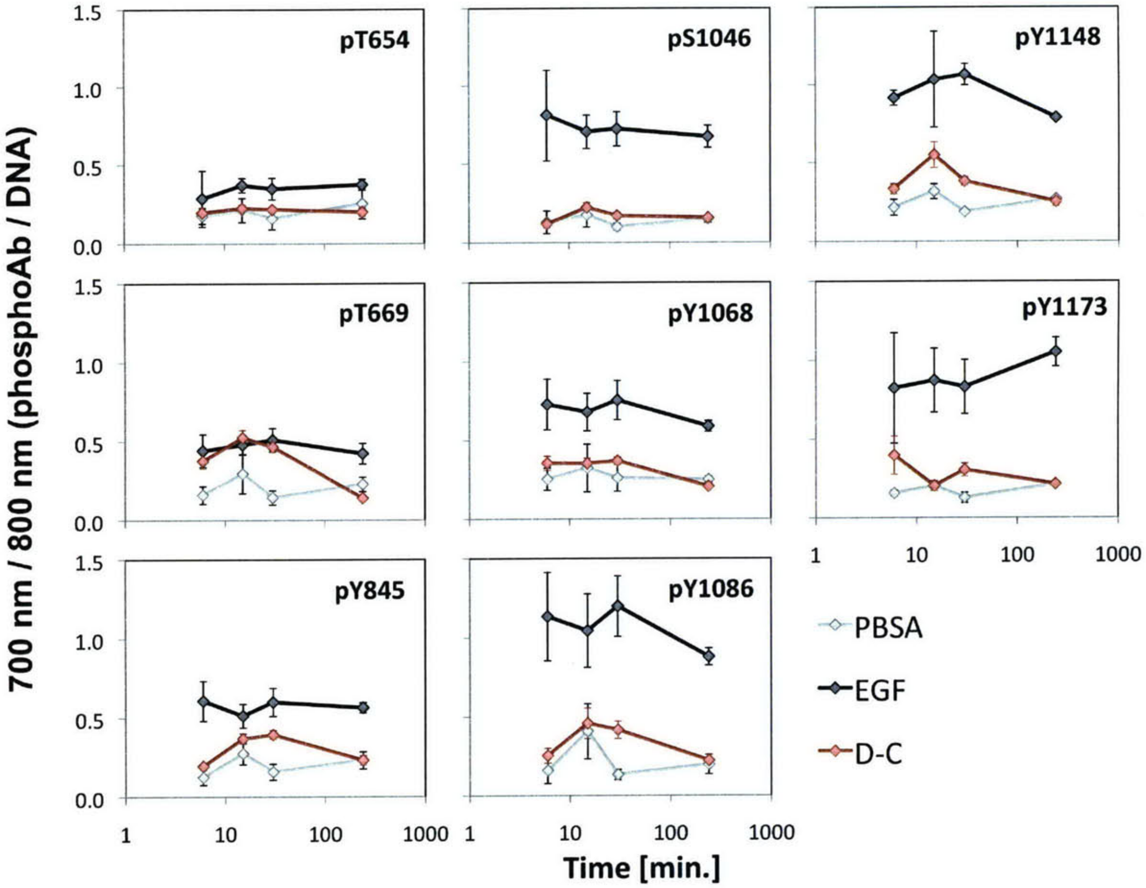


Figure 5.10. *EGFR Agonism.* A431 cells were cultured in 96-well plates, serum starved, and treated with 20 nM agent for 5, 15, 60, or 240 min. Cells were fixed, permeabilized, labeled with rabbit anti-phosho-(S/T/Y) antibody followed by anti-rabbit-800CW and ToPro3 (to stain DNA), and imaged.

Likewise, standard Western blot analysis of cell lysates reveals that heterobivalents do not yield significant phosphorylation of extracellular signal-regulated kinase (ERK1/2) at Y202/Y204 upon 15 minute incubation whereas EGF is activating (Figure 5.11).

This result is corroborated by global phosphorylation analysis of A431 cells upon addition of heterobivalent for 15 or 60 min. Cells were treated with 20 nM agent and phosphorylated tyrosine peptides were analyzed by iTRAQ LC-MS/MS. EGF yields substantially more phosphorylation than heterobivalents or a pair of monovalents (Figure 5.12).

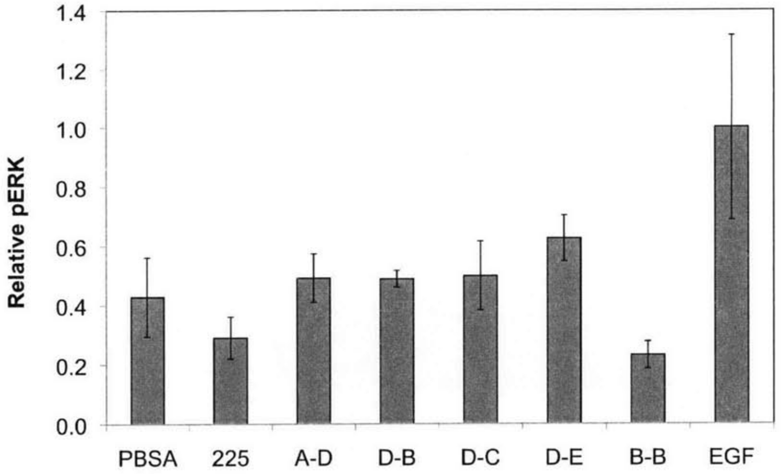


Figure 5.11. *ERK Agonism.* A431 cells were cultured in 24-well plates, serum starved, and treated with 20 nM agent for 15 min. Cell lysates were separated by SDS-PAGE, blotted to nitrocellulose, and labeled with rabbit anti-phosphoERK1/2 Y202/Y204 antibody followed by peroxidase-conjugated anti-rabbit antibody and imaged.

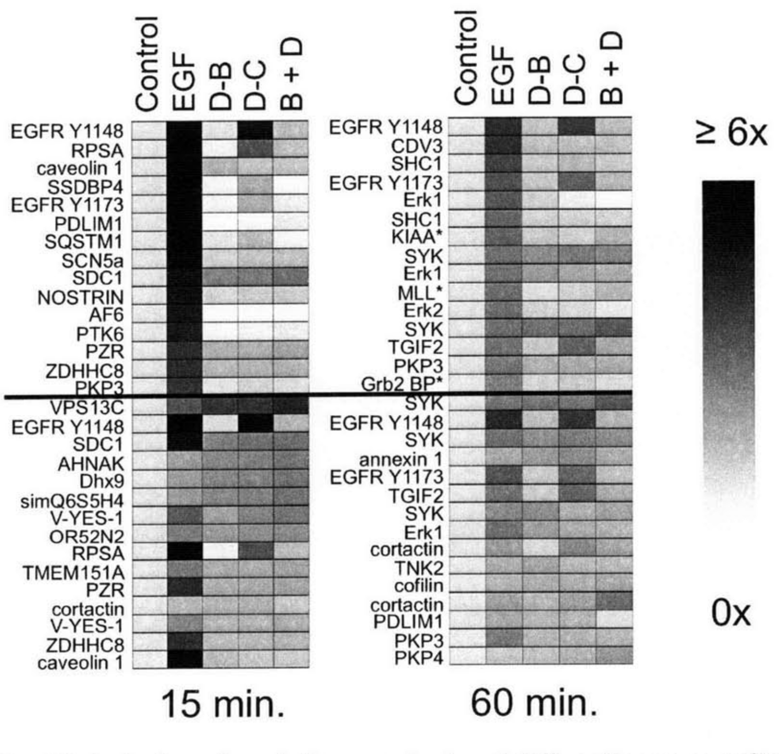


Figure 5.12. *Global phosphorylation analysis.* A431 cells were cultured in 12-well plates, serum starved, and treated with 20 nM agent for 15 or 60 min. Cell lysates were reduced, alkylated, digested, and labeled with iTRAQ isotopic labels. Peptides with phosphorylated tyrosines are isolated by polyclonal antibody affinity chromatography and analyzed by LC-MS/MS. Relative phosphorylation is quantified by comparison of isotopically related peaks. Top portion represents fifteen highest responders to EGF treatment. Lower portion represents fifteen highest responders to heterobivalent treatment.

Collectively, these data demonstrate that select Fn3-Fn3 heterobivalents substantially downregulate EGFR in a manner distinct from EGF and without significant receptor activation.

EGFR Trafficking Model

EGFR trafficking can be examined with a model consisting of four simple mechanisms: synthesis, endocytosis, degradation, and recycling (Figure 5.13).

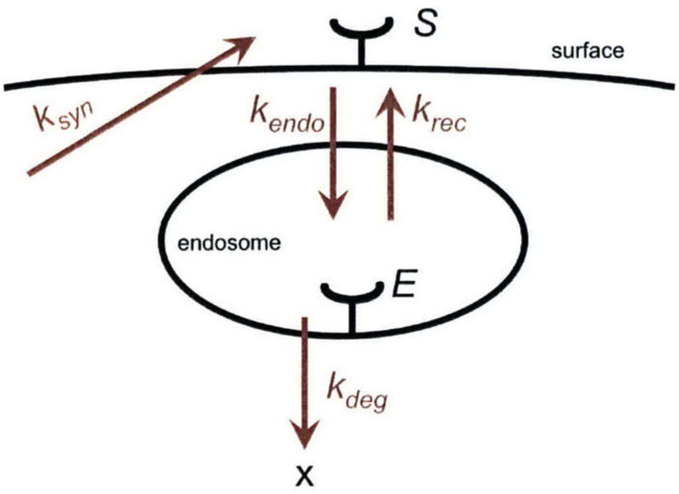


Figure 5.13. *Simple EGFR trafficking model.* Constitutive synthesis produces surface receptor (*S*) at rate k_{syn}. Surface receptor is internalized to endosome (*E*) at rate $k_{endo}S$. Endosomal receptor is degraded at rate $k_{deg}E$ or recycled to the surface at rate $k_{rec}E$.

The behavior of surface receptor (*S*) and endosomal receptor (*E*) are described as follows:

$$\frac{dS}{dt} = k_{syn} - k_{endo}S + k_{rec}E \qquad [5.1]$$

$$\frac{dE}{dt} = k_{endo}S - k_{\deg}E - k_{rec}E \qquad [5.2]$$

The steady state solution is identified by equating the time differentials to zero.

$$E_{ss} = \frac{k_{syn}}{k_{\deg}} \qquad [5.3]$$

$$S_{ss} = \frac{k_{syn}}{k_{endo}}\left(1 + \frac{k_{rec}}{k_{\deg}}\right) \qquad [5.4]$$

Thus, surface receptor can be downregulated via three mechanisms: decreased synthesis, increased endocytosis or decreased recycling fraction (k_{rec}/k_{deg}).

Reduced synthesis reduces fractional receptor levels proportionally:

$$\frac{S_{ss,slow\,syn}}{S_{ss,orig}}=\frac{\frac{k_{syn,slow}}{k_{endo}}\left(1+\frac{k_{rec}}{k_{\deg}}\right)}{\frac{k_{syn,orig}}{k_{endo}}\left(1+\frac{k_{rec}}{k_{\deg}}\right)}=\frac{k_{syn,slow}}{k_{syn,orig}} \qquad [5.5]$$

Enhanced endocytosis reduces fractional receptor levels to the ratio of endocytic rates:

$$\frac{S_{ss,fast}}{S_{ss,slow}}=\frac{\frac{k_{syn}}{k_{endo,fast}}\left(1+\frac{k_{rec}}{k_{\deg}}\right)}{\frac{k_{syn}}{k_{endo,slow}}\left(1+\frac{k_{rec}}{k_{\deg}}\right)}=\frac{k_{endo,slow}}{k_{endo,fast}} \qquad [5.6]$$

For example, if endocytosis with a half-time of one hour, on par with constitutive internalization, is sped to ligand-driven rates with a four minute half-time, the surface receptor level will decrease to 6.7% of original:

$$\frac{S_{ss,fast}}{S_{ss,slow}}=\frac{k_{endo,slow}}{k_{endo,fast}}=\frac{\ln(2)/60\,\mathrm{min}}{\ln(2)/4\,\mathrm{min}}\times 100\%=6.7\% \qquad [5.7]$$

Reduction of the recycling fraction yields the following downregulation:

$$\frac{S_{ss,rec2}}{S_{ss,rec1}}=\frac{\frac{k_{syn}}{k_{endo}}\left[1+\left(\frac{k_{rec}}{k_{\deg}}\right)_2\right]}{\frac{k_{syn}}{k_{endo}}\left[1+\left(\frac{k_{rec}}{k_{\deg}}\right)_1\right]}=\frac{1+\left(\frac{k_{rec}}{k_{\deg}}\right)_2}{1+\left(\frac{k_{rec}}{k_{\deg}}\right)_1} \qquad [5.8]$$

Thus, downregulation via recycling inhibition is limited by the original recycling fraction:

$$\frac{S_{ss,no\,rec}}{S_{ss,rec1}}=\frac{1}{1+\left(\frac{k_{rec}}{k_{\deg}}\right)_1} \qquad [5.9]$$

The kinetics of downregulation in the absence of recycling are determined by solving Equation 5.1:

$$\frac{dS}{dt} = -k_{endo}S + k_{rec}E \rightarrow S = \frac{k_{syn}}{k_{endo}}\left[1 + \left(\frac{k_{rec}}{k_{\deg}}\right)_{orig} e^{-k_{endo}t}\right] \qquad [5.10]$$

The kinetics of downregulation in the absence of recycling, therefore, are solely driven by endocytosis kinetics; *i.e.*, the half-time for downregulation is equal to the endocytic half-time.

The experimental results are consistent with a reduced recycling fraction (either through enhanced degradation or inhibition of recycling). To achieve downregulation to 20% of original, as observed, the recycling fraction would have to be at least 80% originally (see Equation 5.9), which is reasonable. Some heterobivalents may still yield some recycling resulting in reduced downregulation. From a kinetic standpoint, the 0.4-1.4h half-time for downregulation is consistent with the constitutive internalization rates for EGFR (see Equation 5.10).

Enhanced endocytosis could also be considered. A five-fold increase in the rate of endocytosis, which is achievable through agonistic internalization, would yield downregulation to 20%. The dynamic solution to the model differential equations are more complex for enhanced endocytosis than eliminated recycling; yet a numerical solution reveals that a system with an 80% recycling fraction (to be consistent with the reduced recycling fraction hypothesis numbers) and k_{rec} = 11.3 h^{-1} and k_{deg} = 2.8 h^{-1} (derived from French and Lauffenburger[19]) will yield the experimentally observed kinetics for basal endocytosis half-times of 0.6-1.5h, again consistent with constitutive internalization rates for EGFR.

Reduced synthesis can also achieve downregulation to 20% of original levels though the speed of downregulation is much slower (5.2h in the aforementioned system).

Thus, both reduced recycling fraction and enhanced endocytosis rate are quantitatively consistent with the experimental downregulation extent and kinetics. A combination of these mechanisms is also possible.

Efficacy

The ability of monovalent, homobivalent, and heterobivalent constructs to inhibit downstream signaling was examined. The downregulating bivalents A-D, D-B, D-C, and D-E inhibit EGF-induced ERK phosphorylation at tyrosines 202 and/or 204 whereas non-downregulating B-B homobivalent has no effect (Figure 5.14). The monovalent EGF competitor clone D is also antagonistic.

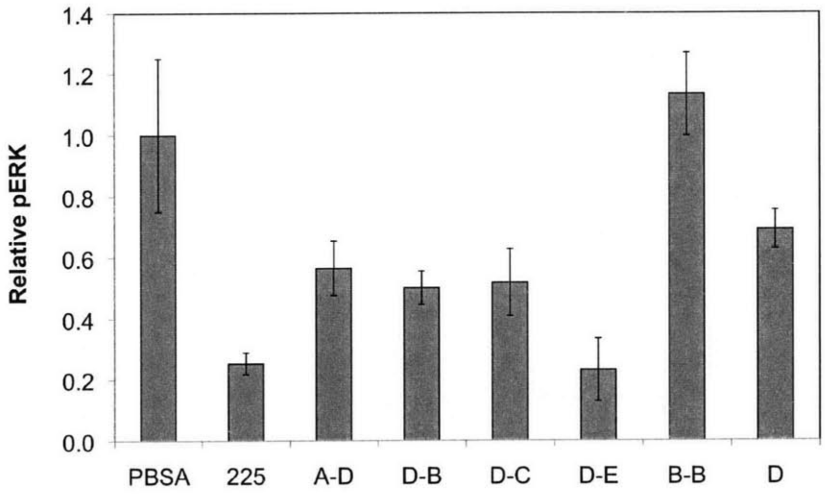


Figure 5.14. *Inhibition of ERK phosphorylation.* A431 cells were cultured in 24-well plates, serum starved, and treated with 20 nM agent for 6h. Cells were then treated with 1 nM EGF for 15 min. Cell lysates were separated by SDS-PAGE, blotted to nitrocellulose, and labeled with rabbit anti-phosphoERK1/2 Y202/Y204 antibody followed by peroxidase-conjugated anti-rabbit antibody and imaged.

Beyond phosphorylation, the effect on cellular output was examined in terms of proliferation and migration. To test cellular output in a challenging tumor-like environment, an autocrine model system was used in which hMEC cells are transfected with a vector for a membrane-bound EGF ligand with an EGF or TGFα cytoplasmic tail (hMEC+ECT or hMEC+TCT[20]). Treatment with downregulating heterobivalent Fn3-Fn3 significantly reduced the number of viable cells at 48h and 96h (Figure 5.15). In addition, combination treatment of 225 antibody and heterobivalent A-D (A and D are not 225 competitive) further reduces cell viability. Of note, clones A and D are not competitive with 225 and thus this combination treatment elicits strong downregulation (Figure 5.16). Likewise, treatment with downregulating heterobivalent strongly reduces cell migration in the autocrine cells as well as parental hMEC cells, and combination treatment further augments this inhibition (Figure 5.17).

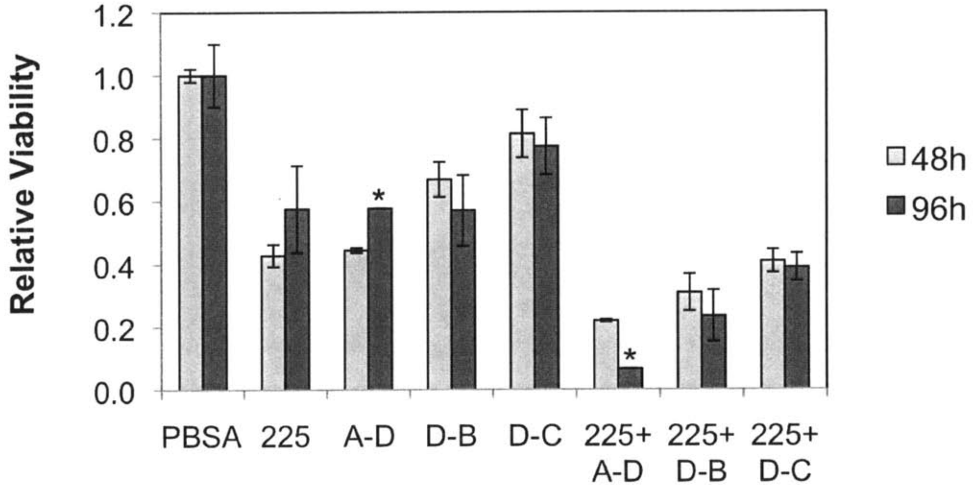


Figure 5.15. *Inhibition of proliferation.* hMEC cells with autocrine EGF signaling were cultured in 96-well plates and treated with 20 nM of the indicated agent(s). Additional ligand is added after 48h. Viability is quantified using AlamarBlue and normalized independently for each time point relative to PBSA-treated cells. Column and error bars represent mean and standard deviation of triplicate samples. * indicates data from a single sample.

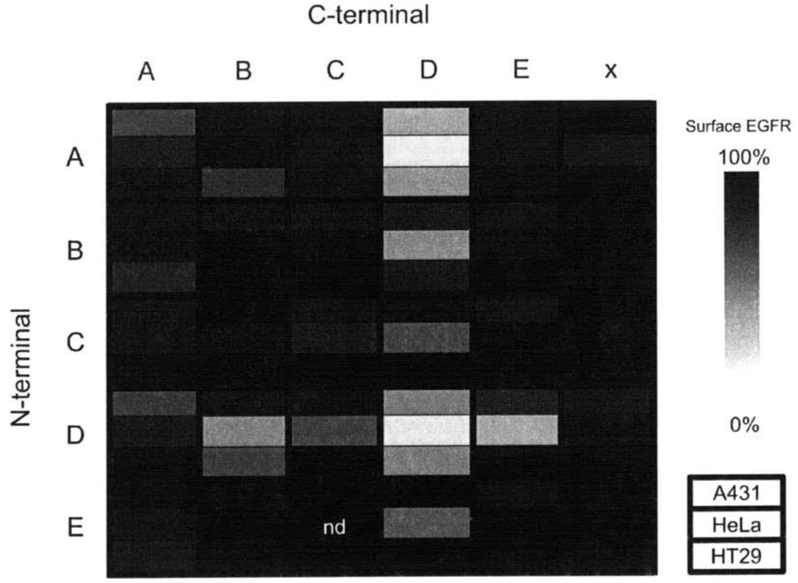


Figure 5.16. *EGFR downregulation with Fn3-Fn3 and 225.* A431, HeLa, and HT29 cells were cultured, serum starved, and treated with 20 nM 225 and 20 nM of the indicated Fn3 or Fn3-Fn3 construct for 6-8h. Surface EGFR was quantified by flow cytometry and is presented on an intensity scale relative to PBSA-treated control with black indicating no downregulation and white indicating complete downregulation. The mean of triplicate samples is presented.

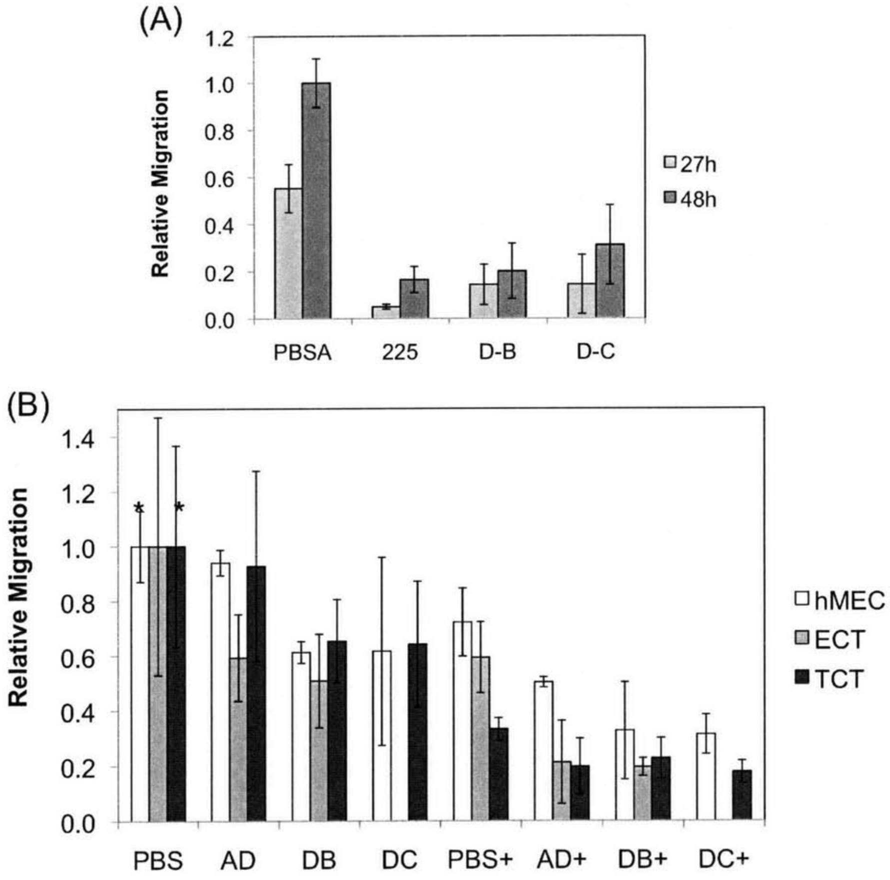


Figure 5.17. *Inhibition of migration.* Cells were cultured in 96-well plates to a confluent monolayer. A 'wound' was scratched into each monolayer to create a void of cells. Cells were treated with 20 nM of the indicated agent(s). Migration was analyzed by microscopy. (A) hMEC cells with autocrine EGF signaling (TCT). (B) hMEC, ECT, and TCT cells. + indicates addition of 225 antibody. * indicates that PBSA 'wound' was completely healed, thus measurable migration was limited. Column and error bars represent mean and standard deviation of triplicate samples.

Delivery

The engineered EGFR binders, both in monovalent and bivalent formats, are effective intracellular delivery agents. Fn3 and Fn3-Fn3 constructs were conjugated to DyLight633 fluorophore via primary amines and incubated with HT29 cells. DyLight633 readily accumulated intracellularly for EGFR binding clones but not for wild-type Fn3 (Figure 5.18A). Biotinylated Fn3 domains loaded onto streptavidin conjugated to AlexaFluor488 and 1.4 nm NanoGold spheres were effectively delivered to EGFR-expressing cells but not EGFR negative cells (Figure 5.18B).

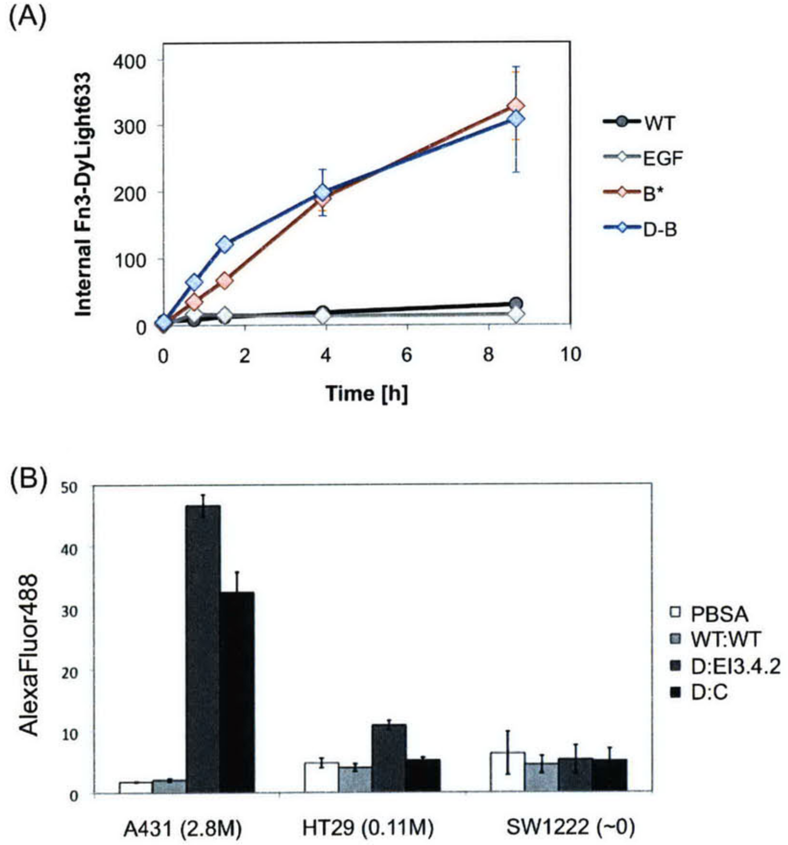


Figure 5.18. *Intracellular delivery.* (A) HT29 cells were cultured in 96-well plates, serum starved, and incubated with 20 nM Fn3-(Fn3)-DyLight633 for the indicated time. Cells were detached with trypsin/EDTA, acid-stripped, and washed. DyLight633 signal was quantified by flow cytometry. (B) Cells were cultured in 96-well plates, serum starved, and incubated in 20 nM biotin-Fn3 :: streptavidin-NanoGold(1.4 nm)-AlexaFluor488 for 12h. Cells were detached with trypsin/EDTA, acid-stripped, and washed. Alexa488 signal was quantified by flow cytometry.

Discussion

The panel of binders should provide useful reagents for a variety of applications. The small size should provide rapid clearance for *in vivo* imaging applications and close proximity of binding site and fluorophore for Förster resonance energy transfer studies. The engineered domains are cysteine-free with primary amines located distal to the presumed binding site with two exceptions: EI1.4.1 contains a cysteine and lysine in the FG loop and clone D contains adjacent cysteines in the FG loop. Thus, the domains are amenable to thiol and amine chemical conjugation to fluorophores, nanoparticles, drug payloads and chemically modified surfaces for drug delivery, diagnostic, and biotechnology applications. The single-domain architecture readily enables protein fusion such as the bivalents discussed herein and immunotoxins (Chris Pirie, unpublished data). The picomolar to low nanomolar binding of these domains is beneficial for most applications. The breadth of epitopes targeted is useful for biophysical studies and dual binding such as for receptor clustering or sandwich immunoassays.

The analysis of the combinations of monovalent and homo- and hetero-bivalent constructs provides a broad data set to assess the stringent criterion for downregulation. As expected, monovalent binding does not reduce EGFR levels. Homobivalents, aside from weak downregulation by D-D, also are ineffective. In fact, strong reduction in EGFR levels is only observed for select heterobivalents of non-competitive clones. Constructs D-B, D-C, and D-E yield the strongest downregulation while A-D, B-D, C-D, and E-D exhibit modest efficacy. Non-competitive heterobivalents including clone D are generally effective except for D-A. Non-competitive heterobivalents including clone A are less consistent. C-A and A-B are weakly effective against all three cell types, A-C and A-E are weakly effective against only two cell types, and B-A and E-A are ineffective. Thus, a combination of non-competitive clones is necessary but not sufficient for strong downregulation. This criterion is consistent with the purported basis for downregulation: receptor clustering. Non-competitive heterobivalent constructs can form receptor clusters because of the ability to bind two heterobivalents to a single receptor thereby propagating receptor linkages whereas homobivalents or competitive heterobivalents can only form two-receptor complexes. Meanwhile, the reduced efficacy of some non-competitive heterobivalents may arise from the inability to simultaneously bind two receptors given the distance and steric constraints of the epitopes targeted and the length and composition of the bivalent linker.

This potential mechanism is also in agreement with the reduced downregulation observed for cells expressing low levels of EGFR as reduced receptor surface density decreases the likelihood of receptor crosslinking. The origin of improved efficacy with shorter linkers is unclear. Perhaps increased conformational flexibility of the Fn3-Fn3 construct reduces the effective local concentration of the unbound Fn3 after single-receptor binding thereby decreasing crosslinking. Alternatively, shorter linkers could increase interaction of clustered receptors though significant agonism is not observed.

The heterobivalents exhibit a response that is grossly different than that elicited by EGF. This is perhaps most clearly demonstrated by the ability of heterobivalents to downregulate EGFR overexpressed in HEK cells, whereas EGF does not downregulate. EGF perhaps fails because of a saturation of the cellular machinery, but regardless the mechanism of downregulation is clearly different for EGF and Fn3-Fn3. Also, multiple receptor mutants, including kinase inactive K721R, are downregulated to the same extent as wild-type receptor. Mutation of neither T669 nor S1046, whose phosphorylation is implicated in receptor internalization[21; 22], nor T654, whose phosphorylation either inhibits ubiquitination or accelerates recycling[23], impacts downregulation. In addition, mutation of Y845, Y1068, Y1148, or Y1173, which are important in the ERK signaling pathway[24-29], has no effect. These results are corroborated by phosphorylation analyses. Of eight key sites studied on EGFR, heterobivalent D-C yielded significantly lower phosphorylation than that by EGF except at T669. Conversely, no phosphorylation is observed at T654, S1046, and Y1068. Y845, Y1086, Y1148, and Y1173 exhibit no agonism at multiple time points and weak phosphorylation at one hour. Moreover, Western blot analysis demonstrates ERK phosphorylation upon treatment with EGF but not upon treatment with any of the heterobivalents tested. Global phosphoproteomic analysis also exhibits substantially more phosphorylation from EGF than D-B, D-C, or a combination of B and D monomers. Thus, unlike EGF, Fn3-Fn3 constructs achieve receptor downregulation without significant receptor agonism.

A simple mathematical model of receptor trafficking indicates that downregulation can be expected to arise from enhanced degradation/recycling ratio, enhanced receptor internalization, or both. The lack of agonism counters the hypothesis of enhanced receptor internalization although endocytosis could be accelerated by weak phosphorylation. Alternatively, the throughput of constitutive internalization could be enhanced via receptor clustering. Yet experimental data suggest that receptor internalization is not sped as monovalent clone B and downregulating D-B exhibit equivalent intracellular accumulation. Moreover, the kinetics of downregulation ($\tau_{1/2}$ = 0.4-1.4h) are comparable to constitutive receptor internalization kinetics. Preliminary measurements of receptor internalization indicate endocytic half-times of 0.3-0.8h (data not shown). Thus, although receptor internalization may be sped slightly, it does not appear to be the dominant source of downregulation. Enhanced degradation could conceivably result from the presence of receptor clusters that either inhibit recycling or drive degradation. In fact, AlexaFluor488-conjugated 225 antibody exhibits reduced recycling in the presence of downregulating heterobivalent A-D as compared to co-treatment with monomer A or non-downregulating C-B (data not shown).

Downregulation decreases the amount of receptor available for ligand binding, receptor homo- and hetero-dimerization, and constitutive activation, thereby decreasing the opportunity for receptor signaling. Downregulation is sufficient to inhibit ERK phosphorylation, a downstream signaling molecule on a pathway that leads to proliferation and migration. Downregulating heterobivalents are shown to inhibit proliferation and migration of a cell line with autocrine signaling, and this inhibitory activity can be augmented by combination treatment with ligand-competitive antibody 225. Further study can elucidate the relative impacts of receptor downregulation and ligand competition as well as the *in vivo* efficacy of the heterobivalent agents.

Materials and Methods

Binder Engineering

EGFR binders were engineered from the NNB, YS, and G4 pooled library comparison as outlined in Chapter 4. EGFR mutant 404SG[Ref. 13] was produced in *Saccharomyces cerevisiae* yeast, purified by metal affinity chromatography and anti-EGFR antibody affinity chromatography, and biotinylated on free amines using the sulfo-NHS biotinylation kit. The Fn3 yeast surface display libraries were pooled, grown in SD-CAA medium at 30°, 250 rpm and display of Fn3 was induced in SG-CAA medium at 30°, 250 rpm. Binders to streptavidin-coated magnetic Dynabeads were removed. One million biotinylated EGFR ectodomains were loaded on each of ten million magnetic beads and incubated with the remaining yeast. Beads were washed once with PBSA at 4° and beads with attached cells were grown for further selection. Remaining sorts were conducted with five million beads coated with one to two million ectodomains. After two sorts, full-length Fn3 clones were selected by FACS using the C-terminal c-myc epitope. Plasmid DNA was zymoprepped from the cells and mutagenized by error-prone PCR of the entire Fn3 gene or the BC, DE, and FG loops. Mutants were transformed into yeast by electroporation with homologous recombination and requisite shuffling of the loop mutants. The lead clones and their mutants were pooled for further cycles of selection and mutagenesis. Three rounds, each consisting of two binding sorts on beads, full-length clone isolation by FACS, and mutagenesis, were performed. Selection stringency was increased by additional washing and elevated temperature. In the fourth round, a single binding sort on magnetic beads was followed by a binding sort by FACS. Cells were incubated in 10 nM biotinylated ectodomain and mouse anti-c-myc antibody followed by fluorescein-conjugated anti-biotin antibody and R-phycoerythrin-conjugated anti-mouse antibody. Cells with the highest fluorescein:R-phycoerythrin ratio were collected. Three additional rounds of sorting and mutagenesis were performed with decreasing ectodomain concentrations during selections. Plasmids from binding populations were zymoprepped and transformed into *E. coli*; transformants were grown, miniprepped, and sequenced.

The relative dominance of E4.2.1 and E4.2.2, as well as very similar mutants, initiated a campaign to identify additional unique clones. Binding populations from rounds two through five were sorted twice for binding to ectodomain in the presence of either ICR10, an antibody that competes with E4.2.2, or 528, an antibody that competes with E4.2.1. Unique clones were identified by sequence analysis.

Fn3 Production

The Fn3 gene was digested with NheI and BamHI and transformed to a pET vector containing a HHHHHHKGSGK-encoding C-terminus. The six histidines enable metal affinity purification, and the pentapeptide provides two additional amines for chemical conjugation. The plasmid was transformed into Rosetta (DE3) *E. coli*, which was grown in LB medium with 100 mg/L kanamycin and 34 mg/L chloramphenicol at 37°. Two hundred μL of overnight culture was added to 100 mL of LB medium, grown to an optical density of 0.2-1.5 units, and induced with 0.5 mM IPTG for 3-24h. Cells were pelleted, resuspended in lysis buffer (50 mM sodium phosphate, pH 8.0, 0.5M NaCl, 5% glycerol, 5 mM CHAPS, 25 mM imidazole, and 1x complete EDTA-free protease inhibitor cocktail), and exposed to four freeze-thaw cycles. The soluble fraction was clarified by centrifugation at 15,000*g* for 10 min. and Fn3 was purified by metal affinity chromatography on TALON resin. Purified Fn3 was buffer exchanged into PBS and biotinylated with NHS-LC-biotin according to the manufacturer's instructions.

An Fn3-linker-Fn3 construct was produced by standard molecular cloning techniques. The resultant vector encodes for Fn3-EIDKPSQ-GSGGGSGGGKGGGGT-Fn3-EIDKPSQ-ELRS-HHHHHH in which the N-terminal Fn3 is bracketed by NheI and BamHI restriction sites and the C-terminal Fn3 is bracketed by KpnI and SacI sites. The reduced linker encodes a GSGT linker. The extended linker is GSGGGSGGGK-GGGSGGGNGGGSGGGGT. Protein was produced as for Fn3.

Affinity Titration

A431 cells were washed in PBSA and incubated with various concentrations of biotinylated Fn3 on ice. The number of cells and sample volumes were selected to ensure excess Fn3 relative to EGFR. For some clones, this criterion necessitates very low cell density, which makes cell collection by centrifugation procedurally difficult. To obviate this difficulty, 'bare' yeast cells are added to the sample to enable effective cell pelleting during centrifugation. Cells were incubated on ice for sufficient time to ensure that the approach to equilibrium was at least 98% complete. Cells were then pelleted, washed with 1 mL PBSA, and incubated in PBSA with 10 mg/L streptavidin-R-phycoerythrin for 10-30 min. Cells were washed and resuspended with PBSA and analyzed by flow cytometry. The minimum and maximum fluorescence and the K_d value were determined by minimizing the sum of squared errors assuming a 1:1 binding interaction.

Epitope Conformational Sensitivity

Yeast were grown and induced to display EGFR ectodomain, incubated at 4° or 80° for 30 min., and chilled on ice for 10 min. Cells were labeled with 40 nM biotinylated Fn3 and 300 nM mouse anti-c-myc antibody followed by streptavidin-R-phycoerythrin and AlexaFluor488-conjugated anti-mouse antibody. Fluorescence was quantified by flow cytometry. Binding (R-phycoerythrin) was normalized to full-length display (AlexaFluor488).

Competition

Yeast displaying EGFR ectodomain or A431 cells were washed and incubated with initial competitor Fn3 or antibody for 30 min. Alternative competitor Fn3, antibody, or AlexaFluor488-conjugated EGF was then added and incubated for 30 min. Cells were washed and secondary reagent was added to detect the alternative competitor: fluorescein-conjugated anti-His antibody, streptavidin-R-phycoerythrin, R-phycoerythrin-conjugated anti-mouse antibody, and fluorescein-conjugated anti-rat antibody for Fn3, biotinylated Fn3, mouse antibodies, and rat ICR10, respectively. Cells were washed and analyzed by flow cytometry. Samples with and without initial competitor were compared to determine competition.

EGFR Fragment Labeling

EGFR ectodomain fragments comprising amino acids 1-176, 294-543, and 302-503 were displayed on the yeast surface.[14] Cells were washed and incubated with 30 nM biotinylated Fn3 and mouse anti-c-myc antibody followed by streptavidin-R-phycoerythrin and AlexaFluor488-conjugated anti-mouse antibody. Cells were washed and analyzed by flow cytometry.

Fine Epitope Mapping

A low mutation library of EGFR ectodomain, produced by Ginger Chao as described,[15] was grown and induced. Yeast were labeled with biotinylated Fn3 and mouse anti-c-myc antibody followed by AlexaFluor647-conjugated streptavidin and AlexaFluor488-conjugated anti-mouse antibody. Cells were washed and analyzed by flow cytometry.

Cells displaying full-length ectodomain (AlexaFluor488^{+}) with reduced Fn3 binding (AlexaFluor647^{weak}) relative to unmutated ectodomain were collected, grown, and induced. Cells were then sorted twice for mutants of reduced binding with maintenance of foldedness as determined by binding to antibodies 199.12 or 225, which are conformationally sensitive.[14] Cells were labeled with biotinylated Fn3 and mouse 199.12 (for clones A, E, and E6.2.10) or mouse 225 (for clone D) anti-EGFR antibody followed by AlexaFluor647-conjugated streptavidin and R-phycoerythrin-conjugated anti-mouse antibody. Cells were washed and analyzed by flow cytometry. Cells displaying folded ectodomain (AlexaFluor488^{+}) with reduced Fn3 binding (AlexaFluor647^{weak}) relative to unmutated ectodomain were collected, grown, and induced. Initial selections for clone C mapping yielded multiple glycine mutants and clones with multiple mutations. To improve the efficiency of folded mutants, analogous sorting was performed using the non-competitive domain III binder clone D for foldedness verification. Biotinylated clones C and D were independently complexed to AlexaFluor488- or AlexaFluor647-conjugated streptavidin and used to label the ectodomain library. Cells that exhibited binding to clone D but reduced clone C binding relative to wild-type ectodomain were collected. Selections for epitope mapping clone B yielded multiple mutants without a consistent location. The full-length ectodomains with reduced clone B binding were sorted for maintenance of clone D binding with a reduction in clone B binding.

Cell Culture

All cells were grown at 37°, 5% CO_2 in a humidified atmosphere. A431 cells were cultured in Dulbecco's modified Eagle medium (DMEM) supplemented with 10% fetal bovine serum (FBS). CHO cells transfected with a vector to express EGFR-green fluorescent protein were cultured in DMEM with 10% FBS, 1% sodium pyruvate, 1% non-essential amino acids, and 0.2 g/L G418. HeLa cells were cultured in Eagle's minimal essential medium with 10% FBS. hMEC cells were cultured in supplemented HuMEC medium. HT29 cells were cultured in McCoy's medium with 10% FBS. U87 cells were cultured in DMEM with 10% FBS, 1% sodium pyruvate, and 1% non-essential amino acids. Cells were detached for subculture or assay use with 0.25% trypsin and 1 mM EDTA. For serum starvation, medium was removed by aspiration, cells were washed with warm PBS, and fresh serum-free medium was added.

Downregulation Assays

Cells were subcultured into 96-well plates, grown for 2 days, and serum starved for 12-18h. Cells were treated with 20 nM Fn3-Fn3 or EGF for the indicated time. Medium was removed by aspiration and cells were washed with PBS, detached with trypsin/EDTA, and placed on ice for the remainder of the assay. Bound Fn3-Fn3 or ligand was removed by 5 min. acid strip with 0.2M acetic acid, 0.5M NaCl. Cells were washed with PBSA and incubated in mouse 225 antibody followed by R-phycoerythrin-conjugated anti-mouse antibody. Cells were washed and analyzed by flow cytometry. Mean fluorescence was normalized to PBSA-treated control samples.

HEK Transfectants

An EGFR expression vector built on the pCDNA3 vector was used as wild-type or modified by site-directed mutagenesis to introduce T654A, T669A, K721R, Y845F, S1045A/S1046A, Y1068F, Y1148F, or Y1173F mutations. Mutation was verified by sequence analysis. HEK cells were grown to 1.2-1.5 million cells per mL and diluted to one million per mL. Miniprepped DNA and polyethyleneimine were independently diluted to 0.05 and 0.1 mg/mL in OptiPro medium and incubated at 22° for 15 min. Equal volumes of DNA and polyethyleneimine were mixed and incubated at 22° for 15 min. 1.2 mL of cells and 48 μL of DNA/polyethyleneimine mixture were added to a 24-well plate and incubated at 37°, 5% CO_2 with shaking for 24h. One hundred μL aliquots of each transfection were transferred to a 96-well plate and grown for 24h. A downregulation assay was performed as described.

In-Cell Western Blot

A431 cells were cultured in 96-well plates, serum starved for 12-24h, and treated with 20 nM Fn3-Fn3 or EGF. Cells were fixed for 10 min. by addition of an equal volume of 4% formaldehyde. Cells were washed and permeabilized with four washes of PBS with 0.1% Triton X100 and blocked in Odyssey blocking buffer for 2h at 22° or overnight at 4°.

Cells were incubated in 10 nM rabbit anti-phospho(S/T/Y) for 2h at 22° or overnight at 4°. Four washes in PBS with 0.1% Tween20 were followed by 33 nM 800CW-conjugated anti-rabbit antibody and 180 nM ToPro3 and four additional washes. Plates were imaged at 700 nm and 800 nm. Antibody signal (800 nm) was normalized to DNA (700 nm) for each well.

Western Blot

A431 cells were cultured in 24-well plates and serum starved for 16h. For agonism assay, cells were treated with 20 nM Fn3-Fn3, antibody, or EGF for 15 min. For antagonism assay, cells were treated with Fn3, Fn3-Fn3, or antibody for 6h followed by 1 nM EGF for 15 min. Medium was removed by aspiration and cells were washed twice with cold PBS and lysed for 5 min. in 50 μL of RIPA buffer with protease and phosphatase inhibitors and EDTA (Pierce). Lysates were clarified by centrifugation at 14,000*g* for 15 min., separated by SDS-PAGE on a 12% BisTris gel, and blotted to nitrocellulose. Blots were blocked in 5% nonfat dry milk and labeled with 1:1000 anti-phosphoERK1/2 Y202/Y204 antibody (Cell Signaling, Danvers, MA) followed by peroxidase-conjugated anti-rabbit antibody. Blots were incubated in SuperSignal West Dura substrate and imaged. Blots were than washed extensively, labeled with rabbit anti-GAPDH antibody followed by peroxidase-conjugated anti-rabbit antibody, incubated with substrate and imaged. PhosphoERK1/2 Y202/Y204 labeling was normalized by GAPDH signal.

Quantitative Phosphoproteomics

A431 cells were cultured in 12-well plates, serum starved for 16h, and treated with 20 nM Fn3-Fn3, Fn3 + Fn3, or EGF for 15 or 60 min. Medium was removed by aspiration and cells were washed with PBS and lysed in 8M urea with 1 mM Na_3VO_4. Phosphoproteomic analysis was performed by Jason Neil of the Forest White lab (MIT). Lysates are digested to form peptides and labeled with iTRAQ reagents. Phosphotyrosine-containing peptides are isolated by immunoprecipitation with a pool of polyclonal anti-phosphotyrosine antibodies and phosphopeptides are enriched by immobilized metal affinity chromatography. Peptides are separated and analyzed by LC-

MS/MS. Peptides are identified using MASCOT and relative abundance is determined by comparison of peak intensities.

Proliferation

hMEC cells transfected with a vector for membrane-bound EGF ligand with a TGFα cytoplasmic tail (hMEC+TCT[20]) were obtained from Doug Lauffenburger (MIT). Eight thousand cells were plated into each well of a 96-well plate and incubated in 100 μL of medium with 20 nM agent for 48h or 96h. For 96h samples, medium was supplemented with fresh agent at 48h. Cell viability was quantified using the AlamarBlue assay (Invitrogen) according the manufacturer's instructions and normalized to PBSA-treated control.

Migration

hMEC, hMEC+ECT, or hMEC+TCT cells were cultured in 96-well plates to confluent monolayers. Wounds were scratched into the monolayer using a pipette tip, and cells were washed with fresh medium and imaged on a Nikon confocal microscope with robotic stage. Cells were treated with 20 nM agent in 100 μL of medium, incubated for 24h or 48h, and imaged at identical fields of view. Migration was quantified as the average reduction in separation across the wound and normalized to PBSA-treated control.

Delivery

Fn3 and Fn3-Fn3 were fluorophore-labeled on primary amines using DyLight633 NHS-ester (Pierce) according to the manufacturer's instructions and extensively desalted. HT29 cells were cultured in 96-well plates, serum starved, and incubated with 20 nM Fn3-(Fn3)-DyLight633 for 0-9h. Cells were detached using trypsin/EDTA, acid stripped in 0.2M acetic acid, 0.5M NaCl for 5 min. and analyzed by flow cytometry.

Biotinylated Fn3 was incubated with streptavidin-NanoGold(1.4 nM)-AlexaFluor488 (Nanoprobes, Yaphank, NY) at a 3:1 Fn3:streptavidin ratio. A431, HT29, and SW1222 cells were cultured in 96-well plates and treated with 20 nM complex for 12h. Cells were detached using trypsin/EDTA, acid stripped in 0.2M acetic acid, 0.5M NaCl, and analyzed by flow cytometry.

References

1. Yarden, Y. & Sliwkowski, M. X. (2001). Untangling the ErbB signalling network. *Nat Rev Mol Cell Biol* **2**, 127-37.
2. Nicholson, R. I., Gee, J. M. & Harper, M. E. (2001). EGFR and cancer prognosis. *Eur J Cancer* **37 Suppl 4**, S9-15.
3. Tateishi, M., Ishida, T., Mitsudomi, T., Kaneko, S. & Sugimachi, K. (1990). Immunohistochemical evidence of autocrine growth factors in adenocarcinoma of the human lung. *Cancer Research* **50**, 7077-80.
4. Lee, J. C., Vivanco, I., Beroukhim, R., Huang, J. H., Feng, W. L., DeBiasi, R. M., Yoshimoto, K., King, J. C., Nghiemphu, P., Yuza, Y., Xu, Q., Greulich, H., Thomas, R. K., Paez, J. G., Peck, T. C., Linhart, D. J., Glatt, K. A., Getz, G., Onofrio, R., Ziaugra, L., Levine, R. L., Gabriel, S., Kawaguchi, T., O'Neill, K., Khan, H., Liau, L. M., Nelson, S. F., Rao, P. N., Mischel, P., Pieper, R. O., Cloughesy, T., Leahy, D. J., Sellers, W. R., Sawyers, C. L., Meyerson, M. & Mellinghoff, I. K. (2006). Epidermal growth factor receptor activation in glioblastoma through novel missense mutations in the extracellular domain. *PLoS Med* **3**, e485.
5. Pedersen, M. W., Meltorn, M., Damstrup, L. & Poulsen, H. S. (2001). The type III epidermal growth factor receptor mutation. Biological significance and potential target for anti-cancer therapy. *Ann Oncol* **12**, 745-60.
6. Sharma, S. V., Bell, D. W., Settleman, J. & Haber, D. A. (2007). Epidermal growth factor receptor mutations in lung cancer. *Nat Rev Cancer* **7**, 169-81.
7. Bonner, J. A., Harari, P. M., Giralt, J., Azarnia, N., Shin, D. M., Cohen, R. B., Jones, C. U., Sur, R., Raben, D., Jassem, J., Ove, R., Kies, M. S., Baselga, J., Youssoufian, H., Amellal, N., Rowinsky, E. K. & Ang, K. K. (2006). Radiotherapy plus cetuximab for squamous-cell carcinoma of the head and neck. *N Engl J Med* **354**, 567-78.
8. Cunningham, D., Humblet, Y., Siena, S., Khayat, D., Bleiberg, H., Santoro, A., Bets, D., Mueser, M., Harstrick, A., Verslype, C., Chau, I. & Van Cutsem, E. (2004). Cetuximab monotherapy and cetuximab plus irinotecan in irinotecan-refractory metastatic colorectal cancer. *N Engl J Med* **351**, 337-45.
9. Messersmith, W. & Hidalgo, M. (2007). Panitumumab, a Monoclonal Anti Epidermal Growth Factor Receptor Antibody in Colorectal Cancer: Another One or the One? *Clinical Cancer Research* **13**, 4664-4666.
10. Perera, R., Narita, Y., Furnari, F. B., Gan, H. K., Murone, C., Ahlkvist, M., Luwor, R., Burgess, A., Stockert, E., Jungbluth, A. A., Old, L. J., Cavenee, W. K., Scott, A. & Johns, T. (2005). Treatment of human tumor xenografts with monoclonal antibody 806 in combination with a prototypical epidermal growth factor receptor-specific antibody generates enhanced antitumor activity. *Clin Cancer Res* **11**, 6390-9.
11. Friedman, L. M., Rinon, A., Schechter, B., Lyass, L., Lavi, S., Bacus, S. S., Sela, M. & Yarden, Y. (2005). Synergistic down-regulation of receptor tyrosine kinases by combinations of mAbs: implications for cancer immunotherapy. *Proc Natl Acad Sci USA* **102**, 1915-20.
12. Ben-Kasus, T., Schechter, B., Lavi, S., Yarden, Y. & Sela, M. (2009). Persistent elimination of ErbB-2/HER2-overexpressing tumors using combinations of monoclonal antibodies: relevance of receptor endocytosis. *Proc Natl Acad Sci USA* **106**, 3294-9.
13. Kim, Y., Bhandari, R., Cochran, J., Kuriyan, J. & Wittrup, K. (2006). Directed evolution of the epidermal growth factor receptor extracellular domain for expression in yeast. *Proteins* **62**, 1026-1035.
14. Cochran, J. R., Kim, Y., Olsen, M. J., Bhandari, R. & Wittrup, K. (2004). Domain-level antibody epitope mapping through yeast surface display of epidermal growth factor receptor fragments. *Journal of Immunological Methods* **287**, 147-58.
15. Chao, G., Cochran, J. R. & Wittrup, K. (2004). Fine epitope mapping of anti-epidermal growth factor receptor antibodies through random mutagenesis and yeast surface display. *Journal of Molecular Biology* **342**, 539-50.

16. Burgess, A., Cho, H. S., Eigenbrot, C., Ferguson, K. M., Garrett, T. P., Leahy, D. J., Lemmon, M. A., Sliwkowski, M. X., Ward, C. W. & Yokoyama, S. (2003). An open-and-shut case? Recent insights into the activation of EGF/ErbB receptors. *Molecular Cell* **12**, 541-52.
17. Ogiso, H., Ishitani, R., Nureki, O., Fukai, S., Yamanaka, M., Kim, J. H., Saito, K., Sakamoto, A., Inoue, M., Shirouzu, M. & Yokoyama, S. (2002). Crystal structure of the complex of human epidermal growth factor and receptor extracellular domains. *Cell* **110**, 775-87.
18. Li, S., Schmitz, K. R., Jeffrey, P. D., Wiltzius, J. J., Kussie, P. & Ferguson, K. M. (2005). Structural basis for inhibition of the epidermal growth factor receptor by cetuximab. *Cancer Cell* **7**, 301-11.
19. French, A. R. & Lauffenburger, D. A. (1996). Intracellular receptor/ligand sorting based on endosomal retention components. *Biotechnol. Bioeng.* **51**, 281-97.
20. Joslin, E. J., Opresko, L. K., Wells, A., Wiley, H. S. & Lauffenburger, D. A. (2007). EGF-receptor-mediated mammary epithelial cell migration is driven by sustained ERK signaling from autocrine stimulation. *J Cell Sci* **120**, 3688-99.
21. Countaway, J. L., Nairn, A. C. & Davis, R. J. (1992). Mechanism of desensitization of the epidermal growth factor receptor protein-tyrosine kinase. *J Biol Chem* **267**, 1129-40.
22. Winograd-Katz, S. E. & Levitzki, A. (2006). Cisplatin induces PKB/Akt activation and p38(MAPK) phosphorylation of the EGF receptor. *Oncogene* **25**, 7381-90.
23. Bao, J., Alroy, I., Waterman, H., Schejter, E. D., Brodie, C., Gruenberg, J. & Yarden, Y. (2000). Threonine phosphorylation diverts internalized epidermal growth factor receptors from a degradative pathway to the recycling endosome. *J Biol Chem* **275**, 26178-86.
24. Amos, S., Martin, P. M., Polar, G. A., Parsons, S. J. & Hussaini, I. M. (2005). Phorbol 12-myristate 13-acetate induces epidermal growth factor receptor transactivation via protein kinase Cdelta/c-Src pathways in glioblastoma cells. *J Biol Chem* **280**, 7729-38.
25. Biscardi, J. S., Maa, M. C., Tice, D. A., Cox, M. E., Leu, T. H. & Parsons, S. J. (1999). c-Src-mediated phosphorylation of the epidermal growth factor receptor on Tyr845 and Tyr1101 is associated with modulation of receptor function. *J Biol Chem* **274**, 8335-43.
26. Downward, J., Waterfield, M. D. & Parker, P. J. (1985). Autophosphorylation and protein kinase C phosphorylation of the epidermal growth factor receptor. Effect on tyrosine kinase activity and ligand binding affinity. *J Biol Chem* **260**, 14538-46.
27. Morandell, S., Stasyk, T., Skvortsov, S., Ascher, S. & Huber, L. A. (2008). Quantitative proteomics and phosphoproteomics reveal novel insights into complexity and dynamics of the EGFR signaling network. *Proteomics* **8**, 4383-401.
28. Wu, W., Graves, L. M., Gill, G. N., Parsons, S. J. & Samet, J. M. (2002). Src-dependent phosphorylation of the epidermal growth factor receptor on tyrosine 845 is required for zinc-induced Ras activation. *J Biol Chem* **277**, 24252-7.
29. Yamauchi, T., Ueki, K., Tobe, K., Tamemoto, H., Sekine, N., Wada, M., Honjo, M., Takahashi, M., Takahashi, T., Hirai, H., Tushima, T., Akanuma, Y., Fujita, T., Komuro, I., Yazaki, Y. & Kadowaki, T. (1997). Tyrosine phosphorylation of the EGF receptor by the kinase Jak2 is induced by growth hormone. *Nature* **390**, 91-6.

6. CLOSING REMARKS

All elements of protein engineering were examined and enhanced to develop a robust platform for engineering stable, specific, high affinity binders based on the Fn3 scaffold.

Library Design

The functional capacity of the scaffold was expanded through inclusion of three-loop diversity and length variability within all three loops. Yet this expansive sequence space requires efficient sampling. Constrained tyrosine/serine diversity is capable of yielding mid-nanomolar binders but is less effective at generating high affinity binding domains. Superior library design is achieved through tailored diversity. Each amino acid is analyzed for its likely functionality: structural stability, binding complementarity, or the possibility of either. The library then biases the designed amino acid distribution to best achieve this functionality (*e.g.*, through wild-type conservation for structural stability or antibody-inspired distribution for binding complementarity). The resultant library, coupled with effective mutagenesis, robustly and swiftly yields high affinity binders. This work is directly applicable to essentially any protein engineering effort including other molecular recognition scaffolds and different functions. Enzyme engineering may particularly benefit because its general necessity of mid-throughput screening limits the number of clones that can be analyzed, placing a premium on functional density.

Further improvements to library design could be achieved through refinement of these approaches. Continued accumulation of isolated clone sequences will identify preferred loop lengths and preferred amino acid distributions for each function. Moreover, the relative effectiveness of each technique for discerning stabilizing sites from binding sites will be elucidated enabling improved tailoring of future libraries. Library size may also be expanded through advances in yeast transformation or increased electroporation replicates. These enhancements may speed binder isolation and identify more and superior binders.

Clone Selections

Highly avid magnetic bead selection facilitates reagent binder depletion, isolation of low affinity binders, and sorting of large populations.[1] This technique complements the previously demonstrated fine affinity discrimination of mid- to high-affinity binders via flow cytometry. A strong relationship between stability and the extent of yeast surface display at elevated temperature enables both stability analysis and the potential for stability selections. Thus, no substantial shortcomings currently exist in clone isolation.

Mutagenesis

Error-prone PCR modeling improves efficiency both procedurally and in terms of sequence space search. Recursive mutagenesis of mid-size populations improves the search breadth, which is beneficial because the ruggedness of the sequence/function landscape means that the best obtainable clone is often not a mutant of the current best clone. Aggressive loop sequence mutation and shuffling through homologous recombination enables substantial sequence changes with reduced loss of function. In parallel, mild gene mutation allows valuable framework mutations and maintains effective loop combinations. As a result, a range of sequence modifications is achieved via a simple protocol. Potential future advances include improved mutagenesis methods to eliminate nucleotide bias in error-prone PCR and elevation of dinucleotide mutations to reduce genetic code bias. Inclusion of naïve loop sequences in shuffling will broaden the sequence space search and prevent suboptimal convergence.

Practical Considerations

The G4 library, or a combination of the NNB, YS, and G4 libraries, should be sorted for binding to antigen immobilized on magnetic beads following depletion of bead binders. After two sorts to yield a mid-size population, full-length clones should be isolated by c-myc$^+$ FACS and mutated using subtle gene mutagenesis and aggressive loop mutagenesis and shuffling, including naïve loops. A new sub-library of unmutated, gene-mutated, and loop-mutated clones is created and the selection cycle continues. Selection stringency can be increased through elevated temperature during sorts, increased washing, and reduced avidity (via shorter induction of Fn3 display and/or reduced antigen immobilization). The selection technique should be switched to FACS once significant labeling (relative to antigen-free control) is observed. This is generally evident in magnetic bead sorts by high yield relative to both earlier sorts and bead depletion sorts of the current population. In apparent contrast to the demonstrated lack of affinity discrimination at high avidity, sequence analysis of multiple engineering campaigns has revealed a relative dominance of a single clone after a few rounds of maturation using beads alone; yet upon switch to FACS selections extensive diversity was revealed. Thus, apparent clone convergence during bead selections should not induce cessation of maturation. Maturation should be continued until the desired phenotype is achieved.

Fibronectin Domains

Fn3 domains are good candidates for any mode of molecular recognition and present unique advantages in several applications. The potential for high throughput binder isolation from the synthetic library, inexpensive bacterial production, and facile chemical conjugation make Fn3 a strong candidate for use in proteomics. Engineered domains demonstrated efficacy when immobilized for affinity purification and fluorophore-conjugated for flow cytometry.

The absence of disulfide bonds, whose presence limits the utility of antibodies in the reducing cytoplasm, makes Fn3 domains intriguing for intracellular applications. Although gene delivery, which is necessary for native production of binders, is still not a clinical reality, intracellular binders are valuable research tools. Moreover, delivery via endocytosis remains a viable possibility especially given the efficient internalization observed for the engineered EGFR binders.

Specific, potent, *in vivo* targeting is another avenue of interest. The small size of Fn3 is advantageous for enhanced clearance of background for *in vivo* imaging and reduced disturbance of conjugated nanoparticles or other payloads. Amino acid composition enables facile thiol and amine chemistry for conjugation and single-domain architecture permits protein fusion. Specifically, binders to CEA, CD276, and EGFR provide tumor-targeting reagents. Though the CD276 and EGFR binders exhibit picomolar affinities, the 2 nM CEA binder may benefit from further affinity maturation for monovalent applications. The CEA and EGFR binders have already been conjugated to fluorophores via primary amines and demonstrate effective cell labeling. In the event of an engineered lysine presenting a primary amine in the binding paratope, mutation may alleviate the problem as was shown for the rabbit IgG binder rI4.5.5; alternatively, introduction of a lone cysteine enables thiol conjugation. The *in vivo* performance of Fn3 domains, both bare and conjugated, must be explored. The comparison to alternative tumor-targeting scaffolds, including antibodies and their domains, is of particular interest.

References

1. Ackerman, M., Levary, D., Tobon, G., Hackel, B., Orcutt, K. D. & Wittrup, K. (2009). Highly avid magnetic bead capture: an efficient selection method for de novo protein engineering utilizing yeast surface display. *Biotechnol Prog* **25**, 774-83.

APPENDIX A. ERROR-PRONE PCR MODELING

Introduction

Many approaches have been suggested for random point mutagenesis of DNA including error-prone PCR (epPCR), chemical mutagenesis, and use of mutator strains.[1] The most effective and widely used approach, epPCR, relies on the inherent errors of DNA amplification that remain uncorrected by DNA polymerases lacking 3'-5' exonuclease activity (*i.e.* non-proofreading enzymes). Early efforts to increase error rate involved the inclusion of manganese and unbalanced nucleotides.[2] Increased mutation can be achieved through additional PCR cycles or use of mutagenic nucleoside analogs.[3] Collectively, these approaches allow an extensive array of mutation from nominally low to as high as 18%.

The aim is for a moderate mutational rate to balance the exponential decline in functional mutants with increased mutation and the ability to accumulate mutations throughout evolution.[4] Yet some propose aggressive mutation, citing a limit to the functional decline[5] and the intention to reach highly mutated sequence space for vastly improved function.[6] However, these claims are weakly substantiated (*e.g.*, a single data point is used to support the limitation in exponential function decline with increased mutation) and apply to 'single-pass' libraries as opposed to evolution. It is likely important to identify double, and perhaps triple, mutants that provide cooperative functionality but the need for many mutations to be identified simultaneously is unproven. Moreover, the probability of identifying several beneficial mutations simultaneously in a single clone is far lower than the probability of accumulating single and double mutations to evolve the multi-mutant clone. Thus, epPCR will be modeled to identify reaction conditions that yield approximately one to four mutations per Fn3 gene (*i.e.*, 1-5% mutation rate for the 101 amino acids).

A second concern with epPCR is that the aforementioned methods have significant mutational bias as unbalanced nucleotides with manganese strongly favor AT→GC, GC→AT, and AT→TA and the nucleoside analogs 8-oxo-dGTP and dPTP strongly favor AT→GC, GC→AT, and AT→CG. Such DNA bias extends to amino acid bias thereby limiting protein diversity by oversampling biased amino acid mutations and undersampling unfavored mutations. The epPCR model will account for the effect of mutational bias under different conditions.

Model Construction

Much of the epPCR model structure is taken from work by Moore and Maranas.[7] The required inputs are (1) the six nucleotide mutation rates (*e.g.* AT→CG). The necessary rates pertain to the mutational frequencies per doubling per nucleotide; however, these rates can be calculated from post-epPCR sequencing data; (2) the number of effective PCR doublings; and (3) the basis for the original sequence either as an explicit sequence or a probability distribution. Note that the mutational rates are considered constant and the imperfect doubling of epPCR is approximated as a number of definite doublings. The model determines a particular initial sequence in accordance with the input basis and generates the resultant mutant amino acid sequences. These sequences are characterized by the number of mutations from the input sequence and analyzed for uniqueness.

Sequence Data Analysis

The probability of a nucleotide being mutated after n doublings is the complement of the probability that it is not mutated, which is the sum of the probability that the nucleotide was never mutated and the probability that it was mutated more than once, resulting in the orignal nucleotide (reversion). For the low frequency mutations of epPCR, the reversion frequency is exceedingly low ($\leq 1\%$ of non-mutation term for $n \leq 25$ doublings and $m \leq 0.1\%$ mutation) and is neglected.

$$\begin{aligned}
\text{p(nt is mutated after n doublings)} &= 1 - \text{p(nt is not mutated after n doublings)} \\
&= 1 - [\text{p(nt is not mutated in any doubling)} + \text{p(nt reverts)}] \\
&\approx 1 - \text{p(nt is not mutateed in any doubling)} \\
&\approx 1 - [\text{p(nt is not mutated in one doubling)}]^{\#\ \text{doublings}} \\
&\approx 1 - [1 - m]^{n} \qquad \text{[A.1]}
\end{aligned}$$

The probability of observing i mutations in a strand of L nucleotides is the product of i mutational probabilities, $L-i$ non-mutational probabilities, and the number of unique combinations of i mutations (denotes as $L \mid i$).

$$\text{p}(i\ \text{mutations}) = (L \mid i)\ [\text{p(nt mutation in } n \text{ doublings)}]^{i}\ [1 - \text{p(nt mutation in } n \text{ doublings)}]^{L-i}$$

$$
\begin{aligned}
&= (L \mid i)\,[1 - (1-m)^n]^i\,\{1 - [1 - (1-m)^n]\}^{L-i} \\
&= (L \mid i)\,[(1-m)^{-n} - 1]^i\,(1-m)^{nL} \qquad \text{[A.2]}
\end{aligned}
$$

The expected number of mutations in a strand of L nucleotides is the sum of the products of the number of mutations and its probability of occurrence.

$$
E(mut) = \sum_{i=0}^{L} i \cdot p(i\ mut) = \sum_{i=0}^{L} i(L \mid i)\left[(1-m)^{-n} - 1\right]^i (1-m)^{nL} \qquad \text{[A.3]}
$$

Not all strands in a PCR mixture are doubled the maximum number of times. Specifically, the number of strands with n doublings in N total doublings (denoted $Z_{n,N}$) is double the number of combinations of n in N.

$$
Z_{n,N} = 2\,(N \mid n) \qquad \text{[A.4]}
$$

The frequency of mutations in a PCR product is the sum of the total mutations divided by the number of nucleotides, the latter of which is the product of the number of strands (2^{N+1}) and the length, L.

$$
\begin{aligned}
m_{overall} &= \frac{1}{L \cdot 2^{N+1}} \sum_{n=0}^{N} Z_{n,N} E(mut.w/n\ doublings) = \frac{1}{L \cdot 2^{N+1}} \sum_{n=0}^{N} \left\{ \frac{2N!}{(N-n)!n!} \sum_{i=1}^{L} \frac{L!}{(i-1)!(L-i)!} \left[(1-m)^{-n} - 1\right]^i (1-m)^{nL} \right\} \\
&= \frac{N!}{L \cdot 2^{N}} \sum_{n=1}^{N} \left\{ \frac{(1-m)^{nL}}{(N-n)!n!} \sum_{i=1}^{L} \frac{L!}{(i-1)!(L-i)!} \left[(1-m)^{-n} - 1\right]^i \right\} \qquad [A.5]
\end{aligned}
$$

Interestingly, despite the apparent non-linearity of the above equation relating m_{overall} and m, the calculated values are nearly perfectly linear (Figure A.1). Thus, though the values of m can be determined precisely for values of m_{overall} by solving the implicit equation above, a linear approximation yields results within 1% for common epPCR conditions. Specifically,

$$
m = m_{overall} / (0.495N + 0.0383) \qquad \text{[A.6]}
$$

Thus, m values can easily be calculated from experimental error-prone PCR data.

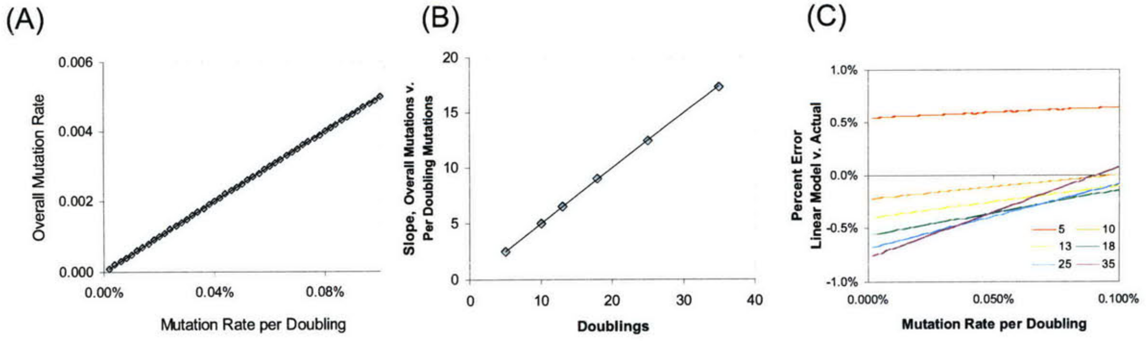


Figure A.1. *Estimation of per doubling nucleotide mutation frequency from overall nucleotide mutation frequency. (A)* The overall mutation frequency (per nucleotide) is plotted versus the per doubling mutation frequency for ten doublings. *(B)* The slope of the best fit line from (A) versus the number of nucleotide doublings. The best fit line is *Slope* = 0.495*doublings* + 0.038. (C) The percentage error for modeling the per doubling mutation rate using the linear relationships demonstrated in (A) and (B).

Mutation Prediction

First, construct a matrix of nucleotide-nucleotide amplification frequencies, M, from post-epPCR sequencing data.

$$M = \begin{bmatrix} m_{A\to A} & m_{A\to C} & m_{A\to G} & m_{A\to T} \\ m_{C\to A} & m_{C\to C} & m_{C\to G} & m_{C\to T} \\ m_{G\to A} & m_{G\to C} & m_{G\to G} & m_{G\to T} \\ m_{T\to A} & m_{T\to C} & m_{T\to G} & m_{T\to T} \end{bmatrix} \qquad [A.7]$$

This per doubling mutation rate matrix can be converted to a mutation rate matrix for n doublings, C^n.

$$C_{i,j}^{n} = \begin{cases} \delta_{ij} & n = 0 \\ M_{ij} & n = 1 \\ \sum\limits_{k=A,C,G,T} M_{kj} C_{ik}^{n-1} & n \geq 2 \end{cases} \qquad [A.8]$$

Not all strands in a PCR mixture are doubled the maximum number of times. Specifically, the number of strands with n doublings in N total doublings (denoted $Z_{n,N}$) is double the number of combinations of n in N.

$$Z_{n,N} = 2\binom{N}{n} \qquad \text{[A.9]}$$

Considering this doubling distribution and the mutation matrix C^n, a total PCR mutation matrix (denoted P^N) can be calculated indicating the probability of nucleotide mutation after N PCR cycles.

$$P_{ij}^N = \frac{1}{2^N}\sum_{n=0}^{N}\binom{N}{n}C_{ij}^n \qquad \text{[A.10]}$$

The final required input is the basis for the DNA used in the epPCR. Any nucleotide distribution (*i.e.*, nucleotide frequency at each position) may be input including defined sequences, libraries of any construction (*e.g.*, triphosphoramidite codons, NNN, NNK, etc), or combinations of the two.

A specific DNA sequence is then generated in accordance with the input distribution. The total PCR mutation frequency matrix, P^N, is used to determine the occurrence probability of each codon (using nucleotide mutation frequencies) at each position. The genetic code converts this codon probability into an amino acid probability. The appropriate number of PCR products (2^N) are then generated in accordance with the amino acid distribution. The generated sequences are analyzed to determine the number of unique sequences and their number of mutations. This data is stored for output. The specific sequence modeling (generation of initial sequence, determination of codon and amino acid probability at each position, and generation of sequences) is repeated multiple times because of the stochastic nature of the model.

Model A.1. *epPCR_Model.m*

```
function [finaldata] = epPCR_Model
clear all;
close all;
tic;

% ************************************************************************
%                              Input variables
% ************************************************************************
% Insert parent seqeunce
start_seq =
'GTTTCTGATGTTCCGAGGGACCTGGAAGTTGTTGCTGCGACCCCCACCAGCCTACTGATCAGCTGGGATGCTCCTGCTGTCACAGTGAGATATTACAGGAT
CACTTACGGAGAAACAGGAGGAAATAGCCCTGTCCAGGAGTTCACTGTGCCTGGGAGCAAGTCTACAGCTACCATCAGCGGCCTTAAACCTGGAGTTGATTA
TACCATCACTGTGTATGCTGTCACTGGCCGTGGAGACAGCCCCGCAAGCAGCAAGCCAATTTCCATTAATTACCGAACAGAAATTGACAAACCATCCCAG';

% Identify reaction conditions
% Matrices may be used for pooled reactions.
% For single reactions, use 1x1 arrays.

% Input mutation matrix identifier (see Mchoose below for indication)
Mmat_index = [2 2 2 20 20 20];
% Input number of cycles
CycleList = [10 15 20 5 10 15];
% Input number of sequences to enumerate
seq_max = 1000;
% Input number of trials
n_trials = 50;

% - - - - - - - - - - - - - - Odds and Ends - - - - - - - - - - - - -
% Calculate number of bases
bases = length(start_seq);
% Calculate number of amino acids
sites = bases/3;
% Prepare n_seq_List
n_seq_List = seq_max*ones(length(CycleList),1);
% Initialize finaldata matrix
finaldata = zeros(sites+2,length(CycleList));

% - - - - - - - - - - - - Loop through conditions - - - - - - - - - -
for iCondition = 1:length(CycleList)

    disp(['Testing condition ', num2str(iCondition)]);

    % Set conditions
    Mchoose = Mmat_index(iCondition);
    n = CycleList(iCondition);
    n_seq = n_seq_List(iCondition);

    % Calculate P matrix using M and C
    P = Calc_Matrix(Mchoose,n);

    ConditionData = zeros(sites+2,1);
    % - - - - - - - - - - - - - Loop through trials - - - - - - - - -
    for iTrials = 1:n_trials

        % Calculate initial codon sequence and AA sequence
        [p_codon1, initial_seq] = Calc_Initial_Seq(sites, start_seq);

        % Calculate final amino acid probability distribution (via codon PD)
        p_aa_cumulative = Calc_AA_PD(sites, P, p_codon1);

        % Create amino acid sequences and analyze uniqueness
        TrialData = CreateAndAnalyze(n_seq, seq_max, sites, p_aa_cumulative, initial_seq);
        ConditionData = ConditionData + TrialData;

    end;
    % __________________________________________________________________

    % Average over trials
    ConditionData = ConditionData ./ n_trials;
    finaldata(:,iCondition) = ConditionData;

end;
% _____________________________________________________________________

% Plot data (seq_cap truncates presented data)
seq_cap = 13;
if length(finaldata) < (seq_cap+2)
    seq_cap = length(finaldata) - 2;
end;

mutlist = linspace(0,seq_cap-1, seq_cap);
ymax = 1.15*max(max(finaldata(2:seq_cap+1,:)));
plot(mutlist,finaldata(2:seq_cap+1,:));
xlabel('Mutations','FontSize',18);
ylabel('Unique Clones','FontSize',18);
axis([0 seq_cap-1 0 ymax]);

LegendForm(:,1) = Mmat_index;
LegendForm(:,2) = CycleList;
```

```
legend(num2str(LegendForm));

toc;

return;

% **********************************************************************
% **********************************************************************
%                     Subfunction Calc_Initial_Seq
%
%               Set initial codon and amino acid sequences
% **********************************************************************
% ~0.005 seconds

function [p_codon1, initial_seq] = Calc_Initial_Seq(sites,start_seq)

% **********************************************************************
%     Set genetic code and triphosphoramidite codons
% **********************************************************************
AA = ['K' 'N' 'K' 'N' 'T' 'T' 'T' 'T' 'R' 'S' 'R' 'S' 'I' 'I' 'M' 'I'...
    'Q' 'H' 'Q' 'H' 'P' 'P' 'P' 'P' 'R' 'R' 'R' 'R' 'L' 'L' 'L' 'L'...
    'E' 'D' 'E' 'D' 'A' 'A' 'A' 'A' 'G' 'G' 'G' 'G' 'V' 'V' 'V' 'V'...
    'Z' 'Y' 'Z' 'Y' 'S' 'S' 'S' 'S' 'Z' 'C' 'W' 'C' 'L' 'F' 'L' 'F'];
TriP = [1 2 8 14 15 19 20 23 28 31 33 34 40 44 48 50 56 58 59 62];

p_codon1 = zeros(64,sites);
initial_seq = zeros(1,sites);
for site_i = 1:sites
    if start_seq(3*site_i-2:3*site_i) == 'tri'
        p_codon1(TriP(ceil(20*rand(1))),site_i) = 1;
    elseif start_seq(3*site_i-2:3*site_i) == 'xxx'
        p_codon1(ceil(64*rand(1)),site_i) = 1;
    else
        for j = 1:3
            if start_seq(3*site_i-3+j) == 'A'
                x(j) = 1;
            elseif start_seq(3*site_i-3+j) == 'C'
                x(j) = 2;
            elseif start_seq(3*site_i-3+j) == 'G'
                x(j) = 3;
            elseif start_seq(3*site_i-3+j) == 'T'
                x(j) = 4;
            else
                disp('Invalid Sequence');
            end;
        end;
        p_codon1(16*(x(1)-1)+4*(x(2)-1)+x(3),site_i) = 1;
    end;
    initial_seq(1,site_i) = AA(find(p_codon1(:,site_i)));
end;

return;
% ______________________________________________________________________
% ______________________________________________________________________

% **********************************************************************
% **********************************************************************
%                         Subfunction Calc_Matrix
%
%                   Calculates P matrix (via M and C)
%
% **********************************************************************
%                             Set mutation matrix
%               Rows 1-4 represent mutations FROM A, C, G, T
%               Columns 1-4 represent mutations TO A, C, G, T
% **********************************************************************
% ~0.0033 seconds (mutation, C, P matrices)
function P = Calc_Matrix(Mchoose,n)

switch Mchoose
    case 0  % 0 uM analogs
M = [0.999907 0 0.000093 0; 0 0.999907 0 0.000093; 0.000093 0 0.999907 ...
    0; 0 0.000093 0 0.999907];
    case 2  % 2 uM analogs
M = [0.997395 0 0.002512 0.000093; 0 0.999535 0 0.000465; 0.000465 0 ...
    0.999535 0; 0.000093 0.002512 0 0.997395];
    case 3  % 2 uM analogs + excess dTTP
M = [0.994179 0 0.004656 0.001165; 0.000107 0.996748 0 0.003145; ...
    0.003145 0 0.996748 0.000107; 0.001165 0.004656 0 0.994179];
    case 20 % 20 uM analogs
M = [0.989858 0.000186 0.009956 0; 0 0.997395 0 0.002605; 0.002605 0 ...
    0.997395 0; 0 0.009956 0.000186 0.989858];
    case 21 % 20 uM analogs + excess dTTP
M = [0.986642 0.000186 0.012100 0.001072; 0.000107 0.994608 0 0.005285; ...
    0.005285 0 0.994608 0.000107; 0.001072 0.012100 0.000186 0.986642];
    case 200    % 200 uM analogs
M = [0.986415 0.005490 0.008095 0; 0 0.997953 0 0.002047; 0.002047 0 ...
    0.997953 0; 0 0.008095 0.005490 0.986415];
    case 201    % 200 uM analogs + excess dTTP
M = [0.983199 0.005490 0.010239 0.001072; 0.000107 0.995166 0 0.004727; ...
    0.004727 0 0.995166 0.000107; 0.001072 0.010239 0.005490 0.983199];
    otherwise
        disp('Invalid entry.  Using 0 uM.');
M = [0.999907 0 0.000093 0; 0 0.999907 0 0.000093; 0.000093 0 0.999907 ...
    0; 0 0.000093 0 0.999907];
end;
% ________________________________________________________________________

% ************************************************************************
%                           Calculate C matrix
%           Rows 4*n+1 to 4*n+4 represent mutations FROM A,C,G,T
%             for a strand that has doubled n times
%           Columns 1-4 represent mutations TO A, C, G, T
% ************************************************************************

C = zeros(4*n+4,4);
for i = 1:4
    C(i,i) = 1;
end;
for i = 1:4
    for j = 1:4
        C(i+4,j) = M(i,j);
    end;
end;
for n_ct = 2:n
    for i = 1:4
        for j = 1:4
            C_val = 0;
            for k = 1:4
                C_val = C_val + M(k,j)*C(4*(n_ct-1)+i,k);
            end;
            C(4*n_ct+i,j) = C_val;
        end;
    end;
end;
% ________________________________________________________________________

% ************************************************************************
%                           Calculate P matrix
%           Rows 1-4 represent mutations from A, C, G, T
%           Columns 1-4 represent mutations to A, C, G, T
% ************************************************************************

P_mat_sum = zeros(4,4);
for n_ct = 0:n
    n_choose = ((factorial(n))/((factorial(n_ct))*...
        (factorial(n-n_ct))));
    P_mat_sum = P_mat_sum + n_choose*C(4*n_ct+1:4*n_ct+4,1:4);
end;
P = (1/(2^n))*P_mat_sum;

return;
% ________________________________________________________________________
% ________________________________________________________________________

% ************************************************************************
% ************************************************************************
%                          Subfunction Calc_AA_PD
%
%            Calculate amino acid probability distribution
%
% ************************************************************************
%            Calculate final codon probability distribution
% ************************************************************************
% ~0.063 seconds

function p_aa_cumulative = Calc_AA_PD(sites, P, p_codon1)

p_codon2 = zeros(64,sites);
for s = 1:sites      % Calculate codon probability at each site
    for c2 = 1:64   % Calculate probability for each codon
        a = 1+floor((c2-1)/16);
        b = 1+floor((c2-16*(a-1)-1)/4);
        c = c2 - 16*(a-1)-4*(b-1);
        for c1 = 1:64
            x = 1+floor((c1-1)/16);
            y = 1+floor((c1-16*(x-1)-1)/4);
            z = c1 - 16*(x-1)-4*(y-1);
            p_codon2(c2,s) = p_codon2(c2,s) + ...
                p_codon1(c1,s)*P(x,a)*P(y,b)*P(z,c);
        end;
    end;
end;
% ________________________________________________________________________

% ************************************************************************
%          Calculate final amino acid probability distribution
% ************************************************************************
% ~0.0033 seconds

% Set  amino acids
AA = ['K' 'N' 'K' 'N' 'T' 'T' 'T' 'T' 'R' 'S' 'R' 'S' 'I' 'I' 'M' 'I'...
    'Q' 'H' 'Q' 'H' 'P' 'P' 'P' 'P' 'R' 'R' 'R' 'R' 'L' 'L' 'L' 'L'...
    'E' 'D' 'E' 'D' 'A' 'A' 'A' 'A' 'G' 'G' 'G' 'G' 'V' 'V' 'V' 'V'...
    'Z' 'Y' 'Z' 'Y' 'S' 'S' 'S' 'S' 'Z' 'C' 'W' 'C' 'L' 'F' 'L' 'F'];
acids = ['A' 'C' 'D' 'E' 'F' 'G' 'H' 'I' 'K' 'L' 'M' 'N' 'P' 'Q' 'R'...
    'S' 'T' 'V' 'W' 'Y' 'Z'];

p_aa2 = zeros(21,sites);
for s = 1:sites     % Calculate amino acid probability at each site
    for aa2 = 1:21  % Calculate probability for each amino acid (+ stop)
        for c = 1:64
            if AA(c) == acids(aa2)
                p_aa2(aa2,s) = p_aa2(aa2,s) + p_codon2(c,s);
            end;
        end;
    end;
    p_aa2(:,s) = (1/sum(p_aa2(:,s)))*p_aa2(:,s);
end;

% Create cumulative amino acid probability
p_aa_cumulative = p_aa2;
for i_cumulative = 2:21
    p_aa_cumulative(i_cumulative,:) = sum(p_aa2(1:i_cumulative,:));
end;
% _______________________________________________________________________

return;
% _______________________________________________________________________
% _______________________________________________________________________

% ***********************************************************************
% ***********************************************************************
%                     Subfunction CreateAndAnalyze
%
%                   Create n_seq amino acid sequences
%
% ***********************************************************************
function TrialData = CreateAndAnalyze(n_seq, seq_max, sites, p_aa_cumulative, initial_seq)

% Prepare vector for final data
TrialData = zeros(sites+2,1);

% Set  amino acids
acids = ['A' 'C' 'D' 'E' 'F' 'G' 'H' 'I' 'K' 'L' 'M' 'N' 'P' 'Q' 'R'...
    'S' 'T' 'V' 'W' 'Y' 'Z'];

% *******************************************************************
%                         Create AA sequences
% *******************************************************************
clear seq_i;
seq_list = char(zeros(n_seq,sites));

% ~0.3 seconds (find takes >90% of time)
for j = 1:n_seq  % Loop to create multiple sequences
    for site_i = 1:sites % Loop to create an amino acid at each site
        val = rand(1);

        i = find(p_aa_cumulative(:,site_i) > val,1);

        if i == 21
            for nz = 1:sites
                seq_list(j,nz) = 'z';
            end;
            TrialData(sites+2,1) = TrialData(sites+2,1) + 1;
            break;
        end;
        seq_list(j,site_i) = acids(i);
    end;
end;

% Fill in random sequences if less than maximum number enumerated
length_i = size(seq_list,1);
if length_i < seq_max
    disp('Maximum number of sequences not enumerated');
    for short_i = 1:(seq_max - length_i)
        seq_list(length_i+short_i,:) = seq_list(ceil(length_i*rand(1)),:);
    end;
end;

% *******************************************************************
%        Analyze PCR sequences for uniqueness and truncations
% *******************************************************************
% ~0.005 seconds

unique_seq = unique(seq_list,'rows');
[r,c] = size(unique_seq);
for i = 1:r
    n_mut = sum(initial_seq(1,:) ~= unique_seq(i,:));
    TrialData(2+n_mut,1) = TrialData(2+n_mut,1) + 1;
end;
num_unique = r;
TrialData(1,1) = TrialData(1,1) + num_unique;

return;
```

Results and Discussion

Mutation Experiment

Wild-type Fn3 was mutated by 15 cycles of error-prone PCR with 0, 2, 20, or 200 μM nucleotide analogs dPTP and 8-oxo-dGTP. Duplicate reactions were performed at each condition and four sequences from each reaction were analyzed to determine the overall mutation rates of each type (*e.g.* AT→CG). Equation A.6 was used to calculate the values for the M matrix (Table A.1).

Table A.1. *Error-prone PCR data.* $m_{overall}$ indicates the mutational frequency for 2880 nucleotides at each condition (labeled by nucleotide analog concentration). m indicates the per-doubling nucleotide mutagenesis rate calculated from the $m_{overall}$ data using Equation A.6.

	0 μM	2 μM	20 μM	200 μM
	$m_{overall}$			
A:T → T:A	0.0000	0.0003	0.0000	0.0000
A:T → C:G	0.0000	0.0000	0.0007	0.0205
A:T → G:C	0.0003	0.0094	0.0372	0.0302
C:G → T:A	0.0003	0.0017	0.0097	0.0076
C:G → G:C	0.0000	0.0000	0.0000	0.0000
C:G → A:T	0.0000	0.0000	0.0000	0.0000
	m			
AA	0.99991	0.99739	0.98986	0.98642
AC	0.00000	0.00000	0.00019	0.00549
AG	0.00009	0.00251	0.00996	0.00810
AT	0.00000	0.00009	0.00000	0.00000
CA	0.00000	0.00000	0.00000	0.00000
CC	0.99991	0.99953	0.99739	0.99795
CG	0.00000	0.00000	0.00000	0.00000
CT	0.00009	0.00047	0.00261	0.00205
GA	0.00009	0.00047	0.00261	0.00205
GC	0.00000	0.00000	0.00000	0.00000
GG	0.99991	0.99953	0.99739	0.99795
GT	0.00000	0.00000	0.00000	0.00000
TA	0.00000	0.00009	0.00000	0.00000
TC	0.00009	0.00251	0.00996	0.00810
TG	0.00000	0.00000	0.00019	0.00549
TT	0.99991	0.99739	0.98986	0.98642

Mutation Model

Historically, epPCR in the Wittrup lab was performed with six conditions: 2 μM analogs for 10 and 20 cycles, 20 μM analogs for 10 and 20 cycles, and 200 μM analogs for 5 and 10 cycles.[8] The epPCR model was executed for an Fn3 library with amino acids 23-30 (DAPAVTVR), 52-56 (GSKST), and 77-86 (GRGDSPASSK) randomized. Fifty trials were computed with 1000 enumerated sequences. This is consistent with directed evolution in which 10,000 lead clones are diversified to create a ten-million member library. In addition, the code was slightly modified to enable collective analysis of sequences pooled from all six conditions. As expected, diverse mutation distributions are achieved depending on the reaction conditions (Figure A.2).

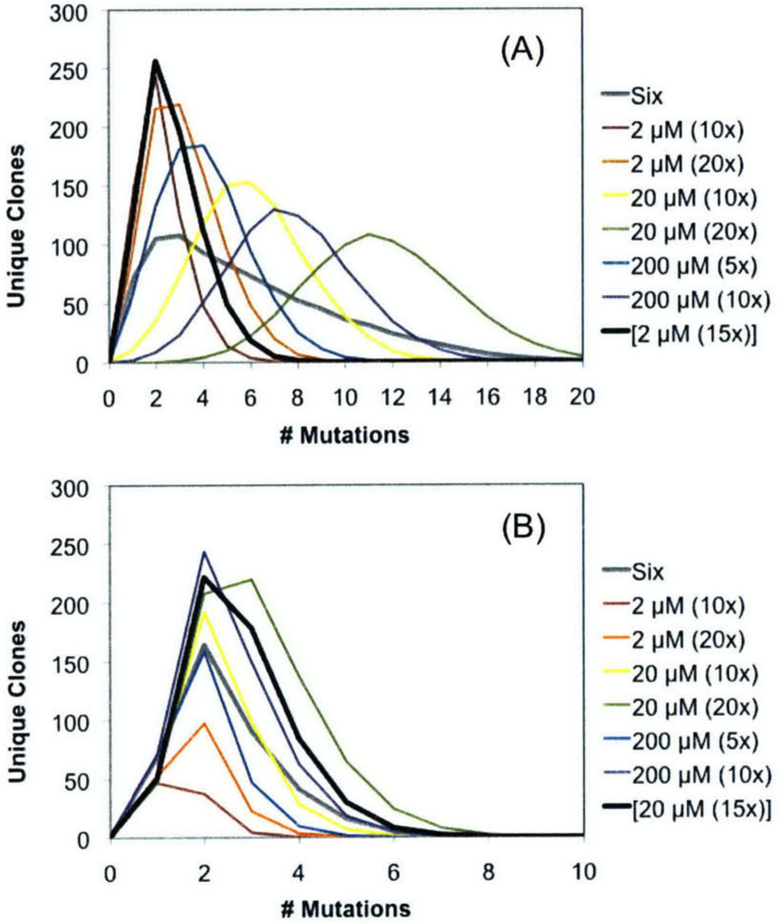


Figure A.2. *Error-prone PCR model results.* Error-prone PCR of an Fn3 library (complete gene (A) or loops alone (B)) was modeled for the six previously used reaction conditions (identified by analog concentration and PCR cycles) as well as a pooled collection of all six conditions. In addition, either a 2 μM, 15-cycle reaction or 20 μM, 15-cycle reaction is tested. The number of unique clones from 1000 sequences is presented.

The resultant mutational distribution from pooling all six reactions includes more high-mutation clones than desired at the expense of mutants with one to five mutations. Thus, for both creation of more desirable mutants as well as improved procedural efficiency, a single preferred reaction condition was identified: 2 μM analogs for 15 cycles. Likewise, the model identified 20 μM analogs for 15 cycles as an effective condition for mutation of the Fn3 loop positions, which entails 23 amino acids at wild-type length. As a result of epPCR modeling, a single reaction can be performed that yield improved mutational distributions relative to the six reaction conditions previously used. The matching cycle numbers enable parallel 15-cycle reactions to be performed for both gene and loop mutagenesis.

This epPCR model was similarly employed to identify an ideal reaction condition for mutagenic libraries of single-chain antibody fragments, interleukin-2, Fc receptor, horseradish peroxidase, and binding peptides.

References

1. Ling, M. M. & Robinson, B. H. (1997). Approaches to DNA mutagenesis: an overview. *Anal Biochem* **254**, 157-78.
2. Cadwell, R. C. & Joyce, G. F. (1992). Randomization of genes by PCR mutagenesis. *PCR Methods Appl* **2**, 28-33.
3. Zaccolo, M., Williams, D. M., Brown, D. M. & Gherardi, E. (1996). An approach to random mutagenesis of DNA using mixtures of triphosphate derivatives of nucleoside analogues. *Journal of Molecular Biology* **255**, 589-603.
4. Shafikhani, S., Siegel, R. A., Ferrari, E. & Schellenberger, V. (1997). Generation of large libraries of random mutants in Bacillus subtilis by PCR-based plasmid multimerization. *BioTechniques* **23**, 304-10.
5. Daugherty, P. S., Chen, G., Iverson, B. L. & Georgiou, G. (2000). Quantitative analysis of the effect of the mutation frequency on the affinity maturation of single chain Fv antibodies. *Proc Natl Acad Sci USA* **97**, 2029-34.
6. Drummond, D. A., Iverson, B. L., Georgiou, G. & Arnold, F. H. (2005). Why high-error-rate random mutagenesis libraries are enriched in functional and improved proteins. *Journal of Molecular Biology* **350**, 806-16.
7. Moore, G. L. & Maranas, C. D. (2000). Modeling DNA mutation and recombination for directed evolution experiments. *Journal of Theoretical Biology* **205**, 483-503.
8. Colby, D. W., Kellogg, B. A., Graff, C. P., Yeung, Y., Swers, J. S. & Wittrup, K. (2004). Engineering antibody affinity by yeast surface display. *Methods in Enzymology* **388**, 348-58.

APPENDIX B. ENGINEERING CEA BINDERS FOR TUMOR TARGETING

Introduction

Carcinoembryonic antigen (CEA) is a 180 kDa glycoprotein in the immunoglobulin superfamily with approximately 50% carbohydrate content.[1] It is expressed in numerous normal tissues, primarily epithelial, with consistently apical localization. However, absence of this polarization as well as occassional upregulation in tumor cells significantly elevates the exposure of CEA to blood and lymphatic vessels rendering CEA as an effective tumor marker. Anti-CEA antibodies have been developed for research and laboratory diagnostics, including an FDA-approved Fab for colorectal cancer imaging.[2] Yet, the reduced size, simple structure, and potentially improved stability of an Fn3 domain could prove beneficial. Moreover, the Fn3 domain could be used for tumor targeting for therapeutic applications.

As outlined in the *Introduction* to Chapter 3, reduced diversity genetic does present an intriguing possibility for improved library design. In particular, serine/tyrosine diversity has been effectively used in antibody libraries.[3-5] This approach enables a substantially more thorough search of theoretical sequence space; yet it achieves this through a selective reduction in sequence space which almost certainly will eliminate many effective clones. The tradeoff of improved efficiency with reduced breadth can be explored through attempts to generate binders from a serine/tyrosine library.

Full diversity and serine/tyrosine diversity libraries are used to engineer binders to CEA. Both libraries yield nanomolar binders thereby demonstrating the success of both designs and generating useful reagents for tumor targeting.

Results and Discussion

Binder Engineering

The NNB and YS yeast surface display Fn3 libraries were independently used to engineer binders to CEA. Four rounds of selection with biotinylated CEA-coated streptavidin magnetic beads and recursive dual mutagenesis yielded binders from both libraries. Sequence analysis revealed that a single clone, NNB C3.2.1 and YS C3.2.3 (Table B.1), dominated each binding population. The serine/tyrosine population was allowed to continue affinity maturation via FACS to investigate the ability to diverge from this constrained code. Sequence analysis throughout affinity maturation revealed multiple point mutations in the BC and FG loops though no strong divergence was observed. After four additional rounds, a dominant clone was identified with three point mutations in the BC loop, an E9G framework mutation, and a framework insert in which VP at positions 4 and 5 were replaced by GTLS.

Table B.1 *CEA binder sequences.* Amino acid sequences of wild-type Fn3 and CEA binders.

Clone	BC	DE	FG	fw
Wild-type	DAPAVTVRY	GSKST	GRGDSPASSK	-
		Initial Top Clones		
NNB C3.2.1	RHVREHY	PRLGR	LGPHV	I34V, E47G, S60G
YS C3.2.3	YYSYSYYYSY	RYRAF	SSSSYSY	V4A
		YS Affinity Maturation		
YS C6.5.1	YYSYSHHYSS	RYRAF	SSSSYSY	VP(4,5)GTLS
YS C6.5.2	YYSYSYHYSS	RYRAF	SSSSYSY	V1A, V4A, D7S, E9G
YS C7.4.3	YYSYSHHYSS	RYRAF	SSSSYSY	VP(4,5)GTLS, E9G
		Targeted Maturation		
Basis	YYSYSYHYSS	RYRAF	SSSSYSY	-
Sh0.3.1	YYPYKYHYHS	RYRAF	ANTNYGY	-
Sh0.3.2	YYPYHYYYSS	RYRAF	ARTPHDY	-
Sh0.3.3	YYPYHYGYSS	RYRAF	TRTPYDY	-
Sh0.3.4	YYPYHYGYSS	RYRAF	RHSPYSY	-
Sh0.3.5	YYPYHYGYSS	RYRAF	ATSPYSY	-
Sh0.3.6	YYPYHYHYNS	RYRAF	ARTDHDY	-

The impacts of framework mutations in clone YS C7.4.3 were tested by reverting each mutation to wild-type both individually and collectively. Yeast were induced to display the clones at elevated temperatures and labeled for binding to biotinylated CEA. Reversion of each mutation to wild-type mildly stabilized Fn3 as indicated by increased display (Figure B.1). The G9E reversion decreases binding to 46±7% of YS C7.4.3 whereas the GTLS(4,5)VP reversion nearly eliminates detectable binding. Thus, in the context of yeast surface display of YS C7.4.3, the insert mutation is critical. Conversely, when produced solubly the framework mutations are not important as YS C7.4.3 and YS

C7.4.3 GTLS(4,5)VP, G9E have equivalent affinities of 1.8 ± 0.4 nM and 2.1 ± 0.6 nM, respectively (Figure B.2).

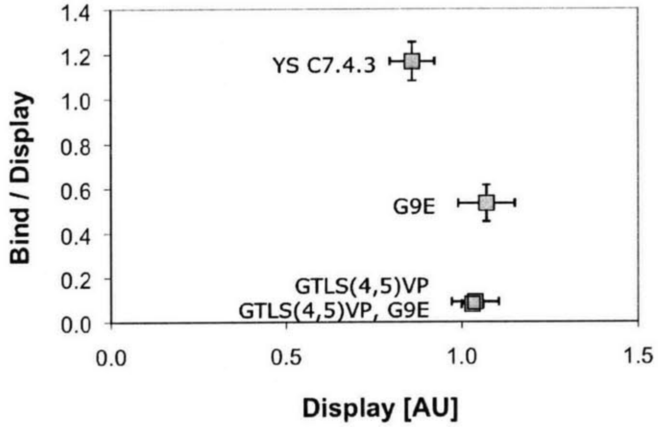


Figure B.1 *YS C7.4.3 framework mutations.* Yeast were induced at 37° to display YS C7.4.3 and framework reversions. Cells were labeled with mouse anti-c-myc antibody and biotinylated CEA followed by AlexaFluor488-conjugated anti-mouse antibody and streptavidin-R-phycoerythrin and analyzed by flow cytometry.

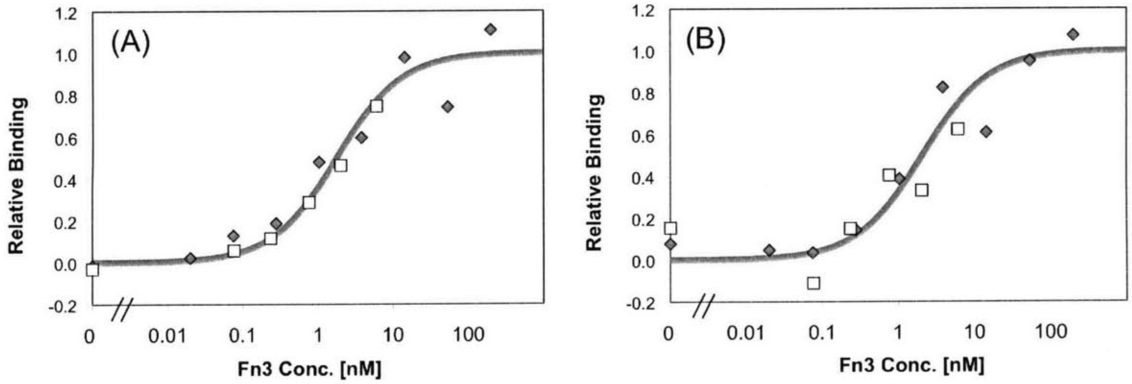


Figure B.2 *Affinity titrations.* YS C7.4.3 (A) and YSC7.4.3 GTLS(4,5)VP, G9E (B) were produced in *E. coli*, purified, and biotinylated. LS174T cells were labeled with the indicated concentration of Fn3, and binding was detected by streptavidin-R-phycoerythrin and flow cytometry.

Synthetic Affinity Maturation

During error-prone PCR affinity maturation, a synthetic affinity maturation scheme was developed in which loops could be independently or collectively matured with a broader search of sequence space. Separate libraries were created for the BC, DE, and FG loops in which each position was diversified to multiple amino acids based on the current top clones YS C6.5.1 and YSC6.5.2. Diversification was initiated at the oligonucleotide synthesis level so more amino acids were achievable than by error-prone PCR; moreover, designed randomization was included in parallel at each position so clones with multiple mutations from the parent sequence were readily achievable, unlike rare "poly-mutants" by error-prone PCR. For the BC library, each position was diversified to six other amino acids. Serines were randomized to D, G, H, N, R, or S; histidines and tyrosines were randomized to C, D, G, H, R, or Y. Two exceptions were S25, which was randomized to P or S because of the structural importance of P25 in wild-type Fn3, and Y29, which was randomized to C, D, F, G, V, or Y because of the structural importance of V29 in wild-type Fn3. The theoretical library size is three million clones. The DE loop was fully randomized to all twenty amino acids at each position, which is a library size of three million clones. In the FG loop, the serines were randomized to A, D, G, H, N, P, R, S, or T and tyrosines were randomized to C, D, F, G, H, L, R, V, or Y, which is a theoretical library of five million clones. Library design is summarized in Table B.2. Three libraries were constructed with a single diversified loop while maintaining the parental binder sequence in the other loops. The libraries were sorted to eliminate non-binders and the successful loops were shuffled to create a library with three diversified loops. This library was sorted thrice and six clones were sequenced.

Table B.2 *Synthetic library design.* The designed amino acids at each position are indicated. *WT* indicates the wild-type amino acid. *Basis* indicates the consensus amino acid in binder sequences. *20* indicates all twenty amino acids.

	BC Loop										*DE Loop*					*FG Loop*						
WT	D	A	P	A	V	-	T	V	R	Y	G	S	K	S	T	G	R	G	D	S	S	K
Basis	Y	Y	S	Y	S	Y	H	Y	S	S	R	Y	R	A	F	S	S	S	S	Y	S	Y
	C	C	P	C	D	C	C	C	D	D	20	20	20	20	20	A	A	A	A	C	A	C
	D	D	S	D	G	D	D	D	G	G						D	D	D	D	D	D	D
	G	G		G	H	G	G	F	H	H						G	G	G	G	F	G	F
	H	H		H	N	H	H	G	N	N						H	H	H	H	G	H	G
	R	R		R	R	R	R	V	R	R						N	N	N	N	H	N	H
	Y	Y		Y	S	Y	Y	Y	S	S						P	P	P	P	L	P	L
																R	R	R	R	R	R	R
																S	S	S	S	V	S	V
																T	T	T	T	Y	T	Y

In the BC loop, the five tyrosines were maintained in all 30 instances. Conversely, serines and histidines were mutated in 18 of 30 instances including all six S25P mutations. The DE loop was fully conserved as parental sequence suggesting that this loop is optimized for affinity, stability, or both. The FG loop is diverse though again tyrosines are well-conserved (10 of 12) whereas serines are not (4 of 30). This result is consistent with the dominant role for tyrosine in molecular recognition with serine playing a largely neutral role.[3]

The VP(4,5)GTLS and E9G framework mutations were added to clone Sh0.3.3, which had the closest match to consensus; yet this clone exhibited slightly weaker binding than YS C7.4.3 (data not shown). Thus, the conserved tyrosines and DE loop seemingly dominate binding while the modified serines have limited effect.

Conclusions

Both the NNB and YS libraries were effective in generation of CEA binders including a clone with 2.1 ± 0.6 nM affinity, which can be pursued for utility in tumor targeting. Yet, while the YS library was effective in generating a nanomolar binder, affinity maturation was largely ineffective as the tyrosine residues appear critical and the serine residues are relatively tolerant of mutation without affecting affinity.

Materials and Methods

Binder Engineering

CEA, purified from colon carcinoma and liver metastases, was purchased from Fitzgerald (Concord, MA). Antigen was biotinylated with NHS-LC-biotin and desalted. The NNB and YS yeast surface display Fn3 libraries were independently grown and induced to display Fn3. Binders to streptavidin-coated magnetic Dynabeads were removed. Streptavidin-coated magnetic Dynabeads were incubated with biotinylated CEA, washed, and incubated with the remaining yeast. The beads were washed with PBSA and the beads with attached cells were grown for further selection. After two magnetic bead sorts, full-length Fn3 clones were selected by fluorescence-activated cell sorting using the C-terminal c-myc epitope for identification of full-length clones. Plasmid DNA was zymoprepped from the cells and mutagenized by error-prone PCR of the entire Fn3 gene or the BC, DE, and FG loops. Mutants were transformed into yeast by electroporation with homologous recombination and requisite shuffling of the loop mutants. The lead clones and their mutants were pooled for further cycles of selection and mutagenesis. Four rounds of engineering were performed. Plasmids from binding populations were zymoprepped and transformed into *E. coli*; transformants were grown, miniprepped, and sequenced.

Engineering of binders from the YS population was continued using FACS selections. Yeast displaying Fn3 were incubated with mouse anti-c-myc antibody and biotinylated CEA followed by fluorophore-conjugated anti-mouse antibody and either fluorophore-conjugated streptavidin or fluorescein-conjugated anti-biotin antibody (to avoid enrichment of streptavidin binders). Clones with the highest CEA binding to c-myc display ratio were collected by flow cytometry.

Mutational Analysis

Wild-type reversion mutants were created by site-directed mutagenesis. Mutants were verified by DNA sequencing. Yeast were transformed with the plasmid clone of interest, grown, and induced at 37°, 250 rpm. Yeast were labeled with mouse anti-c-myc antibody and either 0.5 or 2.5 nM biotinylated CEA followed by AlexaFluor488-conjugated anti-mouse antibody and streptavidin-R-phycoerythrin. The mean fluorescence of the displaying population was quantified by flow cytometry.

The Fn3 gene for YS C7.4.3 and YS C7.4.3 GTLS(4,5)VP, G9E were subcloned into the pET expression vector with a HHHHHHKGSGK C-terminus. Rosetta(DE3) *E. coli* were transformed with plasmid, grown in LB medium, and induced with 0.5 mM IPTG. Cells were lysed by four freeze/thaw cycles and Fn3 was purified from the soluble fraction by metal affinity chromatography with TALON resin. Fn3 was biotinylated with NHS-LC-biotin and desalted.

LS174T cells were cultured in modified Eagle's medium with 10% FBS. Cells were detached with trypsin/EDTA and washed with PBSA. Cells were incubated with various concentrations of biotinylated Fn3 in PBSA followed by streptavidin-R-phycoerythrin and the mean fluorescence was quantified by flow cytometry. The equilibrium dissociation constant was determined by a least squares analysis assuming a 1:1 binding model.

Synthetic Affinity Maturation

Oligonucleotides were designed to produce the library design in Table B.2. Libraries of Fn3 genes containing two parental loops (BC: YYSYSYYYSS; DE: RYRAF; FG: SSSSYSY) and one diverisified loop were created by PCR from the parental sequence using degenerate oligonucleotide primers that overlap the entire diversified loop. Yeast surface display libraries were produced by electroporation of EBY100 yeast with Fn3 gene libraries and linearized pCT-Fn3 vector.

References

1. Hammarström, S. (1999). The carcinoembryonic antigen (CEA) family: structures, suggested functions and expression in normal and malignant tissues. *Semin Cancer Biol* **9**, 67-81.
2. Moffat, F. L., Pinsky, C. M., Hammershaimb, L., Petrelli, N. J., Patt, Y. Z., Whaley, F. S. & Goldenberg, D. M. (1996). Clinical utility of external immunoscintigraphy with the IMMU-4 technetium-99m Fab' antibody fragment in patients undergoing surgery for carcinoma of the colon and rectum: results of a pivotal, phase III trial. The Immunomedics Study Group. *J Clin Oncol* **14**, 2295-305.
3. Fellouse, F., Barthelemy, P. A., Kelley, R. F. & Sidhu, S. (2006). Tyrosine plays a dominant functional role in the paratope of a synthetic antibody derived from a four amino acid code. *Journal of Molecular Biology* **357**, 100-14.
4. Fellouse, F., Li, B., Compaan, D. M., Peden, A. A., Hymowitz, S. G. & Sidhu, S. (2005). Molecular recognition by a binary code. *Journal of Molecular Biology* **348**, 1153-62.
5. Fellouse, F., Wiesmann, C. & Sidhu, S. (2004). Synthetic antibodies from a four-amino-acid code: a dominant role for tyrosine in antigen recognition. *Proc Natl Acad Sci USA* **101**, 12467-72.

APPENDIX C. ENGINEERING CD276 BINDERS FOR TARGETING TUMOR VASCULATURE

Introduction

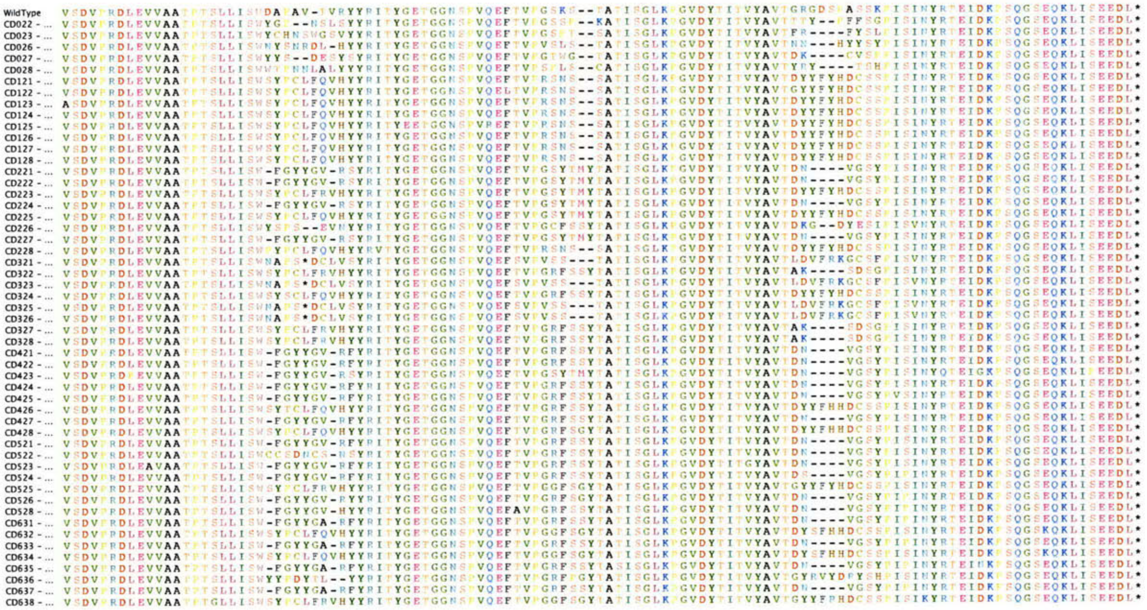
CD276, also known as B7-H3, is a type I membrane protein of the B7 family with a variety of immunological functions.[1] Expression profile analysis identified CD276 as the tumor endothelial marker with the highest differential expression between pathological and physiological angiogenesis.[2] This result was extended to reveal strong differential staining of tumor vasculature relative to normal vessels as well as moderate levels of tumor cell staining. A study of renal cell carcinoma patients revealed CD276 expression on tumor cells in 17% of patients and on tumor vasculature in 95% of patients.[3] Tumor cell or distributed tumor vasculature expression was related to increased risk of death. Also, CD276 is overexpressed in multiple non-small-cell lung cancer cell lines and, in one study, 37% of tumors expressed CD276 and this expression correlated with lymph node metastasis.[4] Thus, though bioligical understanding of this protein is still nascent, it is a potential target for delivery of therapeutic or diagnostic payloads to tumor vasculature. To enable further study of this protein, both for cell biology and as a tumor vascular target, we engineered a panel of Fn3 domains that bind mouse CD276 ectodomain with picomolar affinity.

Results and Discussion

Binder Engineering

The G4 yeast surface display Fn3 library was sorted for binders to mouse CD276 ectodomain. Two rounds of selection with antigen-coated magnetic beads and recursive dual mutagenesis yielded clones capable of binding soluble antigen at mid-nanomolar concentrations (data not shown). Five rounds of FACS selection and mutagenesis yielded binding at picomolar concentrations. Clones from each round of engineering were identified by DNA sequencing (Table C.1). It is noteworthy that a single clone dominates the population enriched by magnetic bead sorting but transition to FACS selections with soluble antigen identifies additional diversity. While a common DE loop motif is dominant from 2.2 onward, diverse BC and FG loop sequences are observed culminating in four unique clones from eight sequences in the final population.

Table C.1. *CD276 binder sequences.* Clones from the c-myc$^+$ population from each round of engineering were sequenced.



WildType	VSDVPRDLEVVAATPTSLLISWDAPAV-TVRYYRITYGETGGNSPVQEFTVPGSKS--TATISGLKPGVDYTITVYAVTGRGDSPASSKPISINYRTEIDKPSQGSEQKLISEEDL*
CD022 - ...	VSDVPRDLEVVAATPTSLLISWYGP--NSLSYYRITYGETGGNSPVQEFTVPGSSP--KATISGLKPGVDYTITVYAVTTTY--PFFSGPISINYRTEIDKPSQGSEQKLISEEDL*
CD023 - ...	VSDVPRDLEVVAATPTSLLISWYCHNSWGSVYYRITYGETGGNSPVQEFTVPGSPT--SATISGLKPGVDYTITVYAVTFR----FYSLPISINYRTEIDKPSQGSEQKLISEEDL*
CD026 - ...	VSDVPRDLEVVAATPTSLLISWNYSNRDL-HYYRITYGETGGNSPVQEFTVPVSLS--TATISGLKPGVDYTITVYAVTNN---HYYSYPISINYRTEIDKPSQGSEQKLISEEDL*
CD027 - ...	VSDVPRDLEVVAATPTSLLISWYYS--DESYSYRITYGETGGNSPVQEFTVPGTWG--TATISGLKPGVDYTITVYAVTDK----CVSSPISINYRTEIDKPSQGSEQKLISEEDL*
CD028 - ...	VSDVPRDLEVVAATPTSLLISWWTPNNLALYYYRITYGETGGNSPVQEFTVPSPLS--CATISGLKPGVDYTITVYAVTYRY----TGHPISINYRTEIDKPSQGSEQKLISEEDL*
CD121 - ...	VSDVPRDLEVVAATPTSLLISWSYPCLFQVHYYRITYGETGGNSPVQEFTVPRSNS--SATISGLKPGVDYTITVYAVTDYYFYHDCSSPISINYRTEIDKPSQGSEQKLISEEDL*
CD122 - ...	VSDVPRDLEVVAATPTSLLISWSYPCLFQVHYYRITYGETGGNSPVQELTVPRSNS--SATISGLKPGVDYTITVYAVTGYYFYHDCSSPISINYRTEIDKPSQGSEQKLISEEDL*
CD123 - ...	ASDVPRDLEVVAATPTSLLISWSYPCLFQVHYYRITYGETGGNSPVQEFTVPRSNS--SATISGLKPGVDYTITVYAVTDYYFYHDCSSPISINYRTEIDKPSQGSEQKLISEEDL*
CD124 - ...	VSDVPRDLEVVAATPTSLLISWSYPCLFQVHYYRITYGETGGNSPVQEFTVPRSNS--SATISGLKPGVDYTITVYAVTDYYFYHDCSSPISINYRTEIDKPSQGSEQKLISEEDL*
CD125 - ...	VSDVPRDLEVVAATPTSLLISWSYPCLFQVHYYRITYEETGGNSPVREFTVPRSNS--SATISGLKPGVDYTITVYAVTDYYFYHDCSSPISINYRTEIDKPSQGSEQKLISEEDL*
CD126 - ...	VSDVPRDLEVVAATPTSLLISWSYPCLFQVHYYRITYGETGGNSPVQEFTVPRSNS--SATISGLKPGVDYTITVYAVTDYYFYHDCSSPISINYRTEIDKPSQGSEQKLISEEDL*
CD127 - ...	VSDVPRDLEVVAATPTSLLISWSYPCLFQVHYYRITYGETGGNSPVQEFTVPRSNS--SATISGLKPGVDYTITVYAVTDYYFYHDCSSPISINYRTEIDKPSQGSEQKLISEEDL*
CD128 - ...	VSDVPRDLEVVAATPTSLLISWSYPCLFQVHYYRITYGETGGNSPVQEFTVPRSNS--SATISGLKPGVDYTITVYAVTDYYFYHDCSSPISINYRTEIDKPSQGSEQKLISEEDL*
CD221 - ...	VSDVPRDLEVVAATPTSLLISW-FGYYGV-RSYRITYGETGGNSPVQEFTVPGSYTMYTATISGLKPGVDYTITVYAVTDN----VGSYPISINYRTEIDKPSQGSEQKLISEEDL*
CD222 - ...	VSDVPRDLEVVAATPTSLLISW-FGYYGV-RSYRITYGETGGNSPVQEFTVPGSYTMYTATISGLKPGVDYTITVYAVTDN----VGSYPISINYRTEIDKPSQGSEQKLISEEDL*
CD223 - ...	VSDVPRDLEVVAATPTSLLISWSYPCLFRVHYYRITYGETGGNSPVQEFTVPGSYTMYTATISGLKPGVDYTITVYAVTDYYFYHDCSSPISINYRTEIDKPSQGSEQKLISEEDL*
CD224 - ...	VSDVPRDLEVVAATPTSLLISW-FGYYGV-RSYRITYGETGGNSPVQEFTVPGSYTMYTATISGLKPGVDYTITVYAVTDN----VGSYPISINYRTEIDKPSQGSEQKLISEEDL*
CD225 - ...	VSDVPRDLEVVAATPTSLLISWSYPCLFQVHYYRITYGETGGNSPVQEFTVPGSYTMYTATISGLKPGVDYTITVYAVTDYYFYHDCSSPISINYRTEIDKPSQGSEQKLISEEDL*
CD226 - ...	VSDVPRDLEVVAATPTSLLISWYSPS--EVNYYRITYGETGGNSPVQEFTVPGCFSSYTATISGLKPGVDYTITVYAVTDKG--DYESIPISVNYRTEIDKPSQGSEQKLISEEDL*
CD227 - ...	VSDVPRDLEVVAATPTSLLISW-FGYYGV-RSYRITYGETGGNSPVQEFTVPGSYTMYTATISGLKPGVDYTITVYAVTDN----VGSYPISINYRTEIDKPSQGSEQKLISEEDL*
CD228 - ...	VSDVPRDLEVVAATPTSLLISWPYPCLFQVHYYRVTYGETGGNSPVQEFTVPRSNS--SATISGLKPGVDYTITVYAVTDYYFYHDCSSPISINYRTEIDKPSQGSEQKLISEEDL*
CD321 - ...	VSDVPRDLEVVAATPTSLLISWNAPS*DCLVSYRITYGETGGNSPVQEFSVPVSS---TATISGLKPGVDYTITVYAVTLDVFRKGCSFPISVNYRTEIDKPSQGSEQKLISEEDL*
CD322 - ...	VSDVPRDLEVVAATPTSLLISWSYPCLFRVHYYRITYGETGGNSPVQEFTVPGRFSSYTATISGLKPGVDYTITVYAVTAK----SDSGPISINYRTEIDKPSQGSEQKLISEEDL*
CD323 - ...	VSDVPRDLEVVAATPTSLLISWNAPS*DCLVSYRITYGETGGNSPVQEFSVPVSS---TATISGLKPGVDYTITVYAVTLDVFRKGCSFPISVNYRTEIDKPSQGSEQKLISEEDL*
CD324 - ...	VSDVPRDLEVVAATPTSLLISWSYSCLFQVHYYRITYGETGGNSPVQEFTVPGRFSSYTATISGLKPGVDYTITVYAVTDYYFYHDCSSPISINYRTEIDKPSQGSEQKLISEEDL*
CD325 - ...	VSDVPRDLEVVAATPTSLLISWNAPS*DCLVSYRITYGETGGNSPVQEFSVPVSS---TATISGLKPGVDYTITVYAVTLDVFRKGCSFPISVNYRTEIDKPSQGSEQKLISEEDL*
CD326 - ...	VSDVPRDLEVVAATPTSLLISWNAPS*DCLVSYRITYGETGGNSPVQEFSVPVSS---TATISGLKPGVDYTITVYAVTLDVFRKGCSFPISVNYRTEIDKPSQGSEQKLISEEDL*
CD327 - ...	VSDVPRDLEVVAATPTSLLISWSYPCLFRVHYYRITYGETGGNSPVQEFTVPGRFSSYTATISGLKPGVDYTITVYAVTAK----SDSGPISINYRTEIDKPSQGSEQKLISEEDL*
CD328 - ...	VSDVPRDLEVVAATPTSLLISWSYPCLFRVHYYRITYGETGGNSPVQEFTVPGRFSSYTATISGLKPGVDYTITVYAVTAK----SDSGPISINYRTEIDKPSQGSEQKLISEEDL*
CD421 - ...	VSDVPRDLEVVAATPTSLLISW-FGYYGV-RFYRITYGETGGNSPVQEFTVPGRFSSYTATISGLKPGVDYTITVYAVTDN----VGSYPISINYRTEIDKPSQGSEQKLISEEDL*
CD422 - ...	VSDVPRDLEVVAATPTSLLISW-FGYYGV-RFYRITYGETGGNSPVQEFTVPGRFSSYTATISGLKPGVDYTITVYAVTDN----VGSYPISINYRTEIDKPSQGSEQKLISEEDL*
CD423 - ...	VSDVPRDPEVVAATPTSLLISW-FGYYGV-RSYRITYGETGGNSPVQEFTVPGSYTMYTATISGLKPGVDYTITVYAVTDN----VGSYPISINYQTEIGKPSQGSEQKLIPEEDL*
CD424 - ...	VSDVPRDLEVVAATPTSLLISW-FGYYGV-RFYRITYGETGGNSPVQEFTVPGRFSSYTATISGLKPGVDYTITVYAVTDN----VGSYPISINYRTEIDKPSQGSEQKLISEEDL*
CD425 - ...	VSDVPRDLEVVAATPTSLLISW-FGYYGV-RFYRITYGETGGNSPVQEFTVPGRFSSYTATISGLKPGVDYTITVYAVTDN----VGSYPVSINYRTEIDKPSQGSEQKLISEEDL*
CD426 - ...	VSDVPRDLEVVAATPTSLLISWSYTCLFQVHYYRITYGETGGNSPVQEFTVPGRFSSYTATISGLKPGVDYTITVYAVTDYYFHHDCSSPISINYRTEIDKPSQGSEQKLISEEDL*
CD427 - ...	VSDVPRDLEVVAATPTSLLISW-FGYYGV-RFYRITYGETGGNSPVQEFTVPGRFSSYTATISGLKPGVDYTITVYAVTDN----VGSYPISINYRTEIDKPSQGSEQKLISEEDL*
CD428 - ...	VSDVPRDLEVVAATPTSLLISWSYPCLFQVHYYRITYGETGGNSPVQEFTVPGRFSGYTATISGLKPGVDYTITVYAVTDYYFHHDCSSPISINYRTEIDKPSQGSEQKLISEEDL*
CD521 - ...	VSDVPRDLEVVAATPTSLLISW-FGYYGV-RFYRITYGETGGNSPVQEFTVPGRFSSYTATISGLKPGVDYTITVYAVTDN----VGSYPISINYRTEIDKPSQGSEQKLISEEDL*
CD522 - ...	VSDVPRDLEVVAATPTSLLISWCCSDNCS-NSYRITYGETGGNSPVQEFTVPGRFSSYTATISGLKPGVDYTITVYAVTDN----VGSYPISINYRTEIDKPSQGSEQKLISEEDL*
CD523 - ...	VSDVPRDLEAVAATPTSLLISW-FGYYGV-RFYRITYGETGGNSPVQEFTVPGRFSSYTATISGLKPGVDYTITGYAVTDN----VGSYPISINYRTEIDKPSQGSEQKLISEEDL*
CD524 - ...	VSDVPRDLEVVAATPTSLLISW-FGYYGV-RFYRITYGETGGNSPVQEFTVPGRFSSYTATISGLKPGVDYTITVYAVTDS----VGSYPIPINYRTEIDKPSQGSEQKLISEEDL*
CD525 - ...	VSDVPRDLEVVAATPTSLLISWSYPCLFRVHYYRITYGETGGNSPVQEFTVPGGFSGYTATISGLKPGVDYTITVYAVTGYYFYHDCSSPISINYRTEIDKPSQGSEQKLISEEDL*
CD526 - ...	VSDVPRDLEVVAATPTSLLISW-FGYYGV-RYYRITYGETGGNSPVQEFTVPGRFSGYTATISGLKPGVDYTITVYAVTDN----VGSYPIPINYRTEIDKPSQGSEQKLISEEDL*
CD528 - ...	VSDVPRDLEVVAATPTSLLISW-FGYYGV-HYYRITYGETGGNSPVQEFAVPGRFSSYTATISGLKPGVDYTITVYAVTDN----VGSYPIPINYRTEIDKPSQGSEQKLISEEDL*
CD631 - ...	VSDVPRDLEVVAATPTSLLISW-FGYYGA-RFYRITYGETGGNSPVQEFTVPGRFSSYTATISGLKPGVDYTITVYAVTDN----VGSYPIPINYRTEIDKPSQGSEQKLISEEDL*
CD632 - ...	VSDVPRDLEVVAATPTSLLISWSYPCLFQVHYYRITYGETGGNSPVQEFTVPGGFSGYTATISGLKPGVDYTITVYAVTDYSFHHDCSSPISINYRTEIDKPSQGSKQKLISEEDL*
CD633 - ...	VSDVPRDLEVVAATPTSLLISW-FGYYGA-RFYRITYGETGGNSPVQEFTVPGRFSSYTATISGLKPGVDYTITVYAVTDN----VGSYPIPINYRTEIDKPSQGSEQKLISEEDL*
CD634 - ...	VSDVPRDLEVVAATPTSLLISWSYPCLFQVHYYRITYGETGGNSPVQEFTVPGGFSGYTATISGLKPGVDYTITVYAVTDYSFHHDCSSPISINYRTEIDKPSQGSKQKLISEEDL*
CD635 - ...	VSDVPRDLEVVAATPTSLLISW-FGYYGA-RFYRITYGETGGNSPVQEFTVPGRFSSYTASISGLKPGVDYTITVYAVTDN----VGSYPIPINYRTEINKPSQGSEQKLISEEDL*
CD636 - ...	VSDVPRDLEVVAATPTSLLISWYYPDYTL--YYRITYGETGGNSPVQEFTVPGRFPGYTATISGLKPGVDYTITVYAVTGYRVYDPYSHPISINYRTEIDKPSQGSEQKLISEEDL*
CD637 - ...	VSDVPRDLEVVAATPTSLLISW-FGYYGA-RFYRITYGETGGNSPVQEFTVPGRFSSYTATISGLKPGVDYTITVYAVTDN----VGSYPIPINYRTEIDKPSQGSEQKLISEEDL*
CD638 - ...	VSDVPRDLEVVAATPTGLLISWSYPCLFRVHYYRITYGETGGNSPVQEFTVPGGFSGYTATISGLKPGVDYTITVYAVTGYYFPHDCSSPISIKYRTEIDKPSQGSEQKLISEEDL*

Clone Analysis

The equilibrium dissociation constant at 22° was determined by titration of soluble biotinylated CD276 for binding to yeast surface displayed CD6.3.1, CD6.3.2, and CD6.3.8. All three clones have picomolar affinity including 1.6 ± 0.9 pM for CD6.3.8 (Figure C.1, Table C.2).

Table C.2. *Top CD276 binders.* Sequences and affinities of clones from round six.

Clone	BC	DE	FG	fw	K_d [pM]
CD6.3.1	FGYYGARF	GRFSSYT	DNVGSY	S89P	54 ± 38
CD6.3.2	SYPCLFQVHY	GGFSGYT	DYSFHHDCSS	E104K	17 ± 14
CD6.3.8	SYPCLFRVHY	GGFSGYT	GYYFRHDCSS	S17G, N91K	1.6 ± 0.9

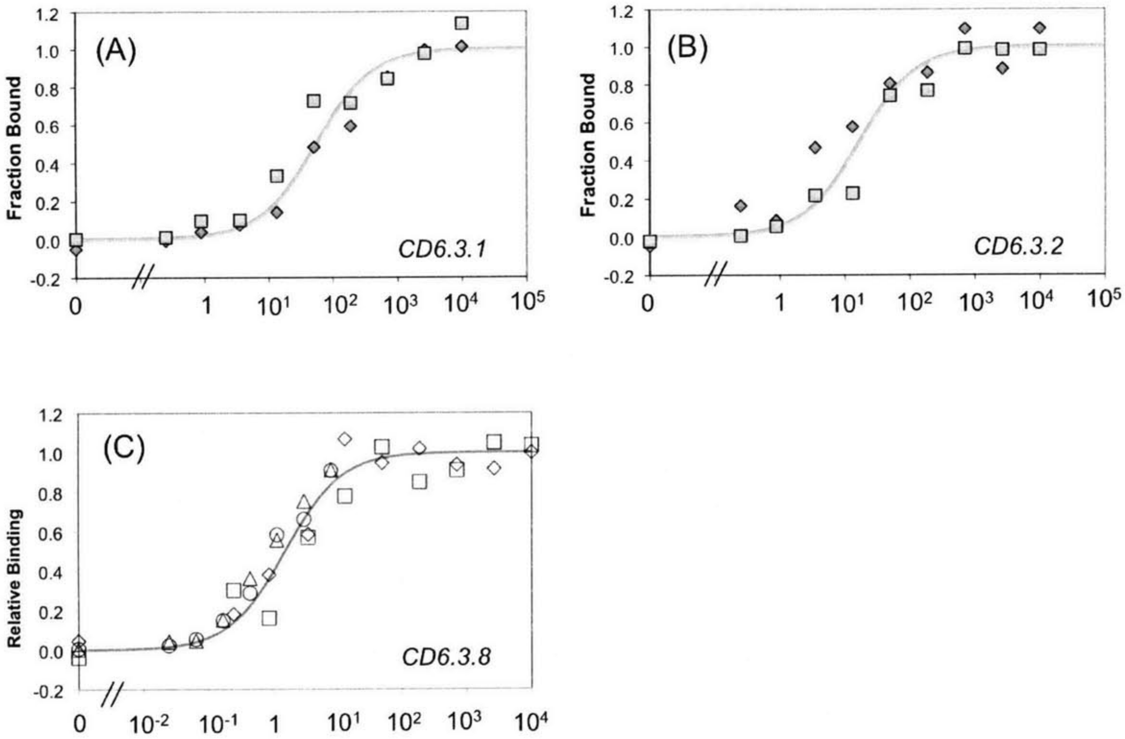


Figure C.1. *Affinity titrations.* Yeast displaying CD6.3.1 (A), CD6.3.2 (B), and CD6.3.8 (C) were incubated with the indicated concentration of biotinylated CD276 at 22° until near equilibrium. Cells were washed and binding was detected with strepavidin-R-phycoerythrin and flow cytometry.

Interestingly, related clones CD6.3.2 and CD6.3.8 contain two cysteine residues that, when modeled in the wild-type Fn3 structure, are in reasonable proximity to form a disulfide bond. In addition, these clones contain proline at position 25, valine at position 29, glycine at position 52, and threonine at position 56, all of which were designed at high frequency in the G4 library for structural bias. Also, the binding loops are devoid of lysine residues enabling chemical conjugation via primary amines, which are present at the N-terminus, K56, and a multi-lysine C-terminal tail, if desired. The engineered high affinity binders should be useful reagents to study CD276 biology and for the development of tumor vasculature targeting agents.

Materials and Methods

Antigen Preparation

Mouse CD276 ectodomain (amino acids 29-244) was purchased from R&D Systems (Minneapolis, MN). It was produced in a mouse myeloma cell line with a human CD33 signal peptide and a C-terminal His_{10} epitope for purification. Protein was resuspended in PBS and biotinylated with NHS-LC-biotin to yield an average of 0.6 biotin molecules per CD276.

Binder Engineering

The G4 yeast surface display Fn3 library was grown and induced to display Fn3. Binders to streptavidin-coated magnetic Dynabeads were removed. Four million streptavidin-coated magnetic Dynabeads were incubated with biotinylated CD276 with one million proteins per bead, washed, and incubated with the remaining yeast. The beads were washed twice with PBSA and the beads with attached cells were grown for further selection. After two magnetic bead sorts, full-length Fn3 clones were selected by fluorescence-activated cell sorting using the C-terminal c-myc epitope for identification of full-length clones. Plasmid DNA was zymoprepped from the cells and mutagenized by error-prone PCR of the entire Fn3 gene or the BC, DE, and FG loops. Mutants were transformed into yeast by electroporation with homologous recombination and requisite shuffling of the loop mutants. The lead clones and their mutants were pooled for further cycles of selection and mutagenesis.

After two rounds of selection and mutagenesis using magnetic beads, FACS selections were performed. Yeast displaying Fn3 were incubated with mouse anti-c-myc antibody and biotinylated CD276 followed by fluorophore-conjugated anti-mouse antibody and either fluorophore-conjugated streptavidin or fluorescein-conjugated anti-biotin antibody (to avoid enrichment of streptavidin binders). Clones with the highest CD276 binding to c-myc display ratio were collected by flow cytometry. Plasmids from binding populations were zymoprepped and transformed into *E. coli*; transformants were grown, miniprepped, and sequenced.

References

1. Chapoval, A. I., Ni, J., Lau, J. S., Wilcox, R. A., Flies, D. B., Liu, D., Dong, H., Sica, G. L., Zhu, G., Tamada, K. & Chen, L. (2001). B7-H3: a costimulatory molecule for T cell activation and IFN-gamma production. *Nat Immunol* **2**, 269-74.
2. Seaman, S., Stevens, J., Yang, M. Y., Logsdon, D., Graff-Cherry, C. & St Croix, B. (2007). Genes that distinguish physiological and pathological angiogenesis. *Cancer Cell* **11**, 539-54.

3. Crispen, P. L., Sheinin, Y., Roth, T. J., Lohse, C. M., Kuntz, S. M., Frigola, X., Thompson, R. H., Boorjian, S. A., Dong, H., Leibovich, B. C., Blute, M. L. & Kwon, E. D. (2008). Tumor cell and tumor vasculature expression of B7-H3 predict survival in clear cell renal cell carcinoma. *Clin Cancer Res* **14**, 5150-7.
4. Sun, Y., Wang, Y., Zhao, J., Gu, M., Giscombe, R., Lefvert, A. K. & Wang, X. (2006). B7-H3 and B7-H4 expression in non-small-cell lung cancer. *Lung Cancer* **53**, 143-51.

APPENDIX D. DNA AND PROTEIN SEQUENCES

pCT-Fn3wt

This construct is used for yeast surface display of Fn3.

Size: 6569 bp
Selectable marker: ampicillin

Aga2--Spacer--FactorXa--HA--Linker--NheI--10Fn3--BamHI--
Myc--2Stop--XhoI--Terminator....

```
ACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGATAATAATGGTTTCTTAGGACGGATCGCTTGCCTGTAACTTACACGCGCCTCGTATCTTTTAATGATGGAATAATTTGGGAATTTACTCTGTGTTTA
TTTATTTTTATGTTTTGTATTTGGATTTTAGAAAGTAAATAAAGAAGGTAGAAGAGTTACGGAATGAAGAAAAAAAAATAAACAAAGGTTTAAAAAATTTCAACAAAAAGCGTACTTTACATATATATTTATTAGACAAGAAA
AGCAGATTAAATAGATATACATTCGATTAACGATAAGTAAAATGTAAAATCACAGGATTTTCGTGTGTGGTCTTCTACACAGACAAGATGAAACAATTCGGCATTAATACCTGAGAGCAGGAAGAGCAAGATAAAAGGTAGTA
TTTGTTGGCGATCCCCCTAGAGTCTTTTACATCTTCGGAAAACAAAAACTATTTTTTCTTTAATTTCTTTTTTTACTTTCTATTTTTAATTTATATATTTATATTAAAAAATTTAAATTATAATTATTTTTATAGCACGTGAT
GAAAAGGACCCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTAT
GAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATC
TCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCT
CAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAA
GGAGCTAACCGCTTTTTTTCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTAT
TAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGG
TCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGCAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTA
ACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAG
ACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACT
GGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTC
GTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCATTGAGAAA
GCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAG
CGTCGATTTTTGTGATGCTCGTCAGGGGGGCCGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCG
TATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGC
ACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTACCTCACTCATTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCCTATGTTGTGTGGAATTGTGAGCGGATAACAATTTCAC
ACAGGAAACAGCTATGACCATGATTACGCCAAGCTCGGAATTAACCCTCACTAAAGGGAACAAAAGCTGGGTACCCGACAGGTTATCAGCAACAACACAGTCATATCCATTCTCAATTAGCTCTACCACAGTGTGTGAACCAA
TGTATCCAGCACCACCTGTAACCAAAACAATTTTAGAAGTACTTTCACTTTGTAACTGAGCTGTCATTTATATTGAATTTTCAAAAATTCTTACTTTTTTTTTGGATGGACGCAAAGAAGTTTAATAATCATATTACATGGCA
TTACCACCATATACATATCCATATACATATCCATATCTAATCTTACTTATATGTTGTGGAAATGTAAAGAGCCCCATTATCTTAGCCTAAAAAAACCTTCTCTTTGGAACTTTCAGTAATACGCTTAACTGCTCATTGCTATA
TTGAAGTACGGATTAGAAGCCGCCGAGCGGGTGACAGCCCTCCGAAGGAAGACTCTCCTCCGTGCGTCCTCGTCTTCACCGGTCGCGTTCCTGAAACGCAGATGTGCCTCGCGCCGCACTGCTCCGAACAATAAAGATTCTAC
AATACTAGCTTTTATGGTTATGAAGAGGAAAAATTGGCAGTAACCTGGCCCCACAAACCTTCAAATGAACGAATCAAATTAACAACCATAGGATGATAATGCGATTAGTTTTTTAGCCTTATTTCTGGGGTAATTAATCAGC
GAAGCGATGATTTTTGATCTATTAACAGATATATAAATGCAAAAACTGCATAACCACTTTAACTAATACTTTCAACATT
TTCGGTTTGTATTACTTCTTATTCAAATGTAATAAAAGTATCAACAAAAAATTGTTAATATACCTCTATACTTTAACGT
CAAGGAGAAAAAACTATAGAATTCTACTTCATACATTTTCAATTAAGATGCAGTTACTTCGCTGTTTTTCAATATTTTC
TGTTATTGCTTCAGTTTTAGCACAGGAACTGACAACTATATGCGAGCAAATCCCCTCACCAACTTTAGAATCGACGCCG
TACTCTTTGTCAACGACTACTATTTTGGCCAACGGGAAGGCAATGCAAGGAGTTTTTGAATATTACAAATCAGTAACGT
TTGTCAGTAATTGCGGTTCTCACCCCTCAACAACTAGCAAAGGCAGCCCCATAAACACACAGTATGTTTTTAAGGACAA
TAGCTCGACGATTGAAGGTAGATACCCATACGACGTTCCAGACTACGCTCTGCAGGCTAGTGGTGGTGGTGGTTCTGGT
GGTGGTGGTTCTGGTGGTGGTGGTTCTGCTAGCGTTTCTGATGTTCCGAGGGACCTGGAAGTTGTTGCTGCGACCCCCA
CCAGCCTACTGATCAGCTGGGATGCTCCTGCTGTCACAGTGAGATATTACAGGATCACTTACGGAGAAACAGGAGGAAA
TAGCCCTGTCCAGGAGTTCACTGTGCCTGGGAGCAAGTCTACAGCTACCATCAGCGGCCTTAAACCTGGAGTTGATTAT
ACCATCACTGTGTATGCTGTCACTGGCCGTGGAGACAGCCCCGCAAGCAGCAAGCCAATTTCCATTAATTACCGAACAG
AAATTGACAAACCATCCCAGGGATCCGAACAAAAGCTTATTTCTGAAGAGGACTTGTAATAGCTCGAGATCTGATAACA
ACAGTGTAGATGTAACAAAATCGACTTTGTTCCCACTGTACTTTTAGCTCGTACAAAATACAATATACTTTTCATTTCT
CCGTAAACAACATGTTTTCCCATGTAATATCCTTTTCTATTTTTCGTTCCGTTACCAACTTTACACATACTTTATATAGc
TATTCACTTCTATACACTAAAAAACTAAGACAATTTTAATTTTGCTGCCTGCCATATTTCAATTTGTTATAAATTCCTATAATTTATCCTATTAGTAGCTAAAAAAAGATGAATGTGAATCGAATCCTAAGAGAATTGAGCTC
CAATTCGCCCTATAGTGAGTCGTATTACAATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCC
GCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGCGCGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCT
TTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAG
TGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTT
CGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACAAAATATTAACGTTTACAATTTCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCAGGCAAGTGCACAAAC
AATACTTAAATAAATACTACTCAGTAATAACCTATTTCTTAGCATTTTTGACGAAATTTGCTATTTTGTTAGAGTCTTTTACACCATTTGTCTCCACACCTCCGCTTACATCAACACCAATAACGCCATTTAATCTAAGCGCA
TCACCAACATTTTCTGGCGTCAGTCCACCAGCTAACATAAAATGTAAGCTTTCGGGGCTCTCTTGCCTTCCAACCCAGTCAGAAATCGAGTTCCAATCCAAAAGTTCACCTGTCCCACCTGCTTCTGAATCAAACAAGGGAAT
AAACGAATGAGGTTTCTGTGAAGCTGCACTGAGTAGTATGTTGCAGTCTTTTGGAAATACGAGTCTTTTAATAACTGGCAAACCGAGGAACTCTTGGTATTCTTGCCACGACTCATCTCCATGCAGTTGGACGATATCAATGC
CGTAATCATTGACCAGAGCCAAAACATCCTCCTTAGGTTGATTACGAAACACGCCAACCAAGTATTTCGGAGTGCCTGAACTATTTTTATATGCTTTTACAAGACTTGAAATTTTCCTTGCAATAACCGGGTCAATTGTTCTC
TTTCTATTGGGCACACATATAATACCCAGCAAGTCAGCATCGGAATCTAGAGCACATTCTGCGGCCTCTGTGCTCTGCAAGCCGCAAACTTTCACCAATGGACCAGAACTACCTGTGAAATTAATAACAGACATACTCCAAGC
TGCCTTTGTGTGCTTAATCACGTATACTCACGTGCTCAATAGTCACCAATGCCCTCCCTCTTGGCCCTCTCCTTTTCTTTTTTCGACCGAATTAATTCTTAATCGGCAAAAAAAGAAAAGCTCCGGATCAAGATTGTACGTAA
GGTGACAAGCTATTTTTCAATAAAGAATATCTTCCACTACTGCCATCTGGCGTCATAACTGCAAAGTACACATATATTACGATGCTGTCTATTAAATGCTTCCTATATTATATATATATAGTAATGTCGTTTATGGTGCACTCTC
AGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCAGAG
GTTTTCACCGTCATCACCGAAACGCGCGA
```

Translation from Aga2p onward:

Aga2p – KDNSST – Xa – HA – $(G_4S)_3$ – AS – $Fn3_{101}$ – GS – c-myc

```
MQLLRCFSIFSVIASVLAQELTTICEQIPSPTLESTPYSLSTTTILANGKAMQGVFEYY
KSVTFVSNCGSHPSTTSKGSPINTQYVFKDNSSTIEGRYPYDVPDYALQASGGGGSGGG
GSGGGGSASVSDVPRDLEVVAATPTSLLISWDAPAVTVRYYRITYGETGGNSPVQEFTV
PGSKSTATISGLKPGVDYTITVYAVTGRGDSPASSKPISINYRTEIDKPSQGSEQKLIS
EEDLZZ
```

pCT-Fn3-Gene

This construct is used for homologous recombination with mutagenized Fn3 genes. The linker-Fn3 portion of pCT-Fn3wt is replaced by PstI-NdeI-BamHI to enable triple digestion to produce linearized vector.

Size: 6227 bp
Selectable marker: ampicillin

Aga2--Spacer--FactorXa--HA--PstI--spacer--NdeI--spacer--
BamHI--Myc--2Stop--XhoI--Terminator....

```
ACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGATAATAATGGTTTCTTAGGACGGATCGCTTGCCTGTAACTTACACGCGCCTCGTATCTTTTAATGATGGAATAATTTGGGAATTTACTCTGTGTTTA
TTTATTTTTATGTTTTGTATTTGGATTTTAGAAAGTAAATAAAGAAGGTAGAAGAGTTACGGAATGAAGAAAAAAAAATAAACAAAGGTTTAAAAAATTTCAACAAAAAGCGTACTTTACATATATATTTATTAGACAAGAAA
AGCAGATTAAATAGATATACATTCGATTAACGATAAGTAAAATGTAAAATCACAGGATTTTCGTGTGTGGTCTTCTACACAGACAAGATGAAACAATTCGGCATTAATACCTGAGAGCAGGAAGAGCAAGATAAAAGGTAGTA
TTTGTTGGCGATCCCCCTAGAGTCTTTTACATCTTCGGAAAACAAAAACTATTTTTTCTTTAATTTCTTTTTTTACTTTCTATTTTTAATTTATATATTTATATTAAAAAATTTAAATTATAATTATTTTTATAGCACGTGAT
GAAAAGGACCCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTAT
GAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATC
TCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCT
CAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAA
GGAGCTAACCGCTTTTTTTCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTAT
TAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGG
TCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGCAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTA
ACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAG
ACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACT
GGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTC
GTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCATTGAGAAA
GCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGGAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAG
CGTCGATTTTTGTGATGCTCGTCAGGGGGGCCGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCG
TATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGC
ACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTACCTCACTCATTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCCTATGTTGTGTGGAATTGTGAGCGGATAACAATTTCAC
ACAGGAAACAGCTATGACCATGATTACGCCAAGCTCGGAATTAACCCTCACTAAAGGGAACAAAAGCTGGGTACCCGACAGGTTATCAGCAACAACACAGTCATATCCATTCTCAATTAGCTCTACCACAGTGTGTGAACCAA
TGTATCCAGCACCACCTGTAACCAAAACAATTTTAGAAGTACTTTCACTTTGTAACTGAGCTGTCATTTATATTGAATTTTCAAAAATTCTTACTTTTTTTTTGGATGGACGCAAAGAAGTTTAATAATCATATTACATGGCA
TTACCACCATATACATATCCATATACATATCCATATCTAATCTTACTTATATGTTGTGGAAATGTAAAGAGCCCCATTATCTTAGCCTAAAAAAACCTTCTCTTTGGAACTTTCAGTAATACGCTTAACTGCTCATTGCTATA
TTGAAGTACGGATTAGAAGCCGCCGAGCGGGTGACAGCCCTCCGAAGGAAGACTCTCCTCCGTGCGTCCTCGTCTTCACCGGTCGCGTTCCTGAAACGCAGATGTGCCTCGCGCCGCACTGCTCCGAACAATAAAGATTCTAC
AATACTAGCTTTTATGGTTATGAAGAGGAAAAATTGGCAGTAACCTGGCCCCACAAACCTTCAAATGAACGAATCAAATTAACAACCATAGGATGATAATGCGATTAGTTTTTTAGCCTTATTTCTGGGGTAATTAATCAGC
GAAGCGATGATTTTTGATCTATTAACAGATATATAAATGCAAAAACTGCATAACCACTTTAACTAATACTTTCAACATT
TTCGGTTTGTATTACTTCTTATTCAAATGTAATAAAAGTATCAACAAAAAATTGTTAATATACCTCTATACTTTAACGT
CAAGGAGAAAAAACTATAGAATTCTACTTCATACATTTTCAATTAAGATGCAGTTACTTCGCTGTTTTTCAATATTTTC
TGTTATTGCTTCAGTTTTAGCACAGGAACTGACAACTATATGCGAGCAAATCCCCTCACCAACTTTAGAATCGACGCCG
TACTCTTTGTCAACGACTACTATTTTGGCCAACGGGAAGGCAATGCAAGGAGTTTTTGAATATTACAAATCAGTAACGT
TTGTCAGTAATTGCGGTTCTCACCCCTCAACAACTAGCAAAGGCAGCCCCATAAACACACAGTATGTTTTTAAGGACAA
TAGCTCGACGATTGAAGGTAGATACCCATACGACGTTCCAGACTACGCTCTGCAGCAGTCGCATATGACTGAGGGATCC
GAACAAAAGCTTATTTCTGAAGAGGACTTGTAATAGCTCGAGATCTGATAACAACAGTGTAGATGTAACAAAATCGACT
TTGTTCCCACTGTACTTTTAGCTCGTACAAAATACAATATACTTTTCATTTCTCCGTAAACAACATGTTTTCCCATGTA
ATATCCTTTTCTATTTTTCGTTCCGTTACCAACTTTACACATACTTTATATAGCTATTCACTTCTATACACTAAAAAACT
AAGACAATTTTAATTTTGCTGCCTGCCATATTTCAATTTGTTATAAATTCCTATAATTTATCCTATTAGTAGCTAAAAAAAGATGAATGTGAATCGAATCCTAAGAGAATTGAGCTCCAATTCGCCCTATAGTGAGTCGTATT
ACAATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCCTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTG
CGCAGCCTGAATGGCGAATGGCGCGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGC
CACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTT
TTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTG
ATTTAACAAAAATTTAACGCGAATTTTAACAAAATATTAACGTTTACAATTTCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCAGGCAAGTGCACAAACAATACTTAAATAAATACTACTCAGTA
ATAACCTATTTCTTAGCATTTTTGACGAAATTTGCTATTTTGTTAGAGTCTTTTACACCATTTGTCTCCACACCTCCGCTTACATCAACACCAATAACGCCATTTAATCTAAGCGCATCACCAACATTTTCTGGCGTCAGTCC
ACCAGCTAACATAAAATGTAAGCTTTCGGGGCTCTCTTGCCTTCCAACCCAGTCAGAAATCGAGTTCCAATCCAAAAGTTCACCTGTCCCACCTGCTTCTGAATCAAACAAGGGAATAAACGAATGAGGTTTCTGTGAAGCTG
CACTGAGTAGTATGTTGCAGTCTTTTGGAAATACGAGTCTTTTAATAACTGGCAAACCGAGGAACTCTTGGTATTCTTGCCACGACTCATCTCCATGCAGTTGGACGATATCAATGCCGTAATCATTGACCAGAGCCAAAACA
TCCTCCTTAGGTTGATTACGAAACACGCCAACCAAGTATTTCGGAGTGCCTGAACTATTTTTATATGCTTTTACAAGACTTGAAATTTTCCTTGCAATAACCGGGTCAATTGTTCTCTTTCTATTGGGCACACATATAATACC
CAGCAAGTCAGCATCGGAATCTAGAGCACATTCTGCGGCCTCTGTGCTCTGCAAGCCGCAAACTTTCACCAATGGACCAGAACTACCTGTGAAATTAATAACAGACATACTCCAAGCTGCCTTTGTGTGCTTAATCACGTATA
CTCACGTGCTCAATAGTCACCAATGCCCTCCCTCTTGGCCCTCTCCTTTTCTTTTTTCGACCGAATTAATTCTTAATCGGCAAAAAAAGAAAAGCTCCGGATCAAGATTGTACGTAAGGTGACAAGCTATTTTTCAATAAAGA
ATATCTTCCACTACTGCCATCTGGCGTCATAACTGCAAAGTACACATATATTACGATGCTGTCTATTAAATGCTTCCTATATTATATATATAGTAATGTCGTTTATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATA
GTTAAGCCAGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCG
CGA
```

pCT-Fn3-Loop

This construct is used for homologous recombination with mutagenized Fn3 loops. A fragment of the Fn3 gene from the BC loop through the FG loop is replaced by NcoI/SmaI/NdeI to enable triple digestion to produce linearized vector.

Size: 6407 bp
Selectable marker: ampicillin

Aga2--Spacer--FactorXa--HA--Linker--NheI--Fn3$_{1-22}$--NcoI--
spacer--SmaI--spacer--NdeI--Fn3$_{87-101}$--BamHI--Myc--2Stop--
XhoI--Terminator....

```
ACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGATAATAATGGTTTCTTAGGACGGATCGCTTGCCTGTAACTTACACGCGCCTCGTATCTTTTAATGATGGAATAATTTGGGAATTTACTCTGTGTTTA
TTTATTTTTATGTTTTGTATTTGGATTTTAGAAAGTAAATAAAGAAGGTAGAAGAGTTACGGAATGAAGAAAAAAAAATAAACAAAGGTTTAAAAAATTTCAACAAAAAGCGTACTTTACATATATATTTATTAGACAAGAAA
AGCAGATTAAATAGATATACATTCGATTAACGATAAGTAAAATGTAAAATCACAGGATTTTCGTGTGTGGTCTTCTACACAGACAAGATGAAACAATTCGGCATTAATACCTGAGAGCAGGAAGAGCAAGATAAAAGGTAGTA
TTTGTTGGCGATCCCCCTAGAGTCTTTTACATCTTCGGAAAACAAAAACTATTTTTTCTTTAATTTCTTTTTTTACTTTCTATTTTTAATTTATATATTTATATTAAAAAATTTAAATTATAATTATTTTTATAGCACGTGAT
GAAAAGGACCCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTAT
GAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATC
TCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCT
CAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAA
GGAGCTAACCGCTTTTTTTCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTAT
TAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGG
TCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGCAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTA
ACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAG
ACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACT
GGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTC
GTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCATTGAGAAA
GCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGGAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAG
CGTCGATTTTTGTGATGCTCGTCAGGGGGGCCGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCG
TATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGC
ACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTACCTCACTCATTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCCTATGTTGTGTGGAATTGTGAGCGGATAACAATTTCAC
ACAGGAAACAGCTATGACCATGATTACGCCAAGCTCGGAATTAACCCTCACTAAAGGGAACAAAAGCTGGGTACCCGACAGGTTATCAGCAACAACACAGTCATATCCATTCTCAATTAGCTCTACCACAGTGTGTGAACCAA
TGTATCCAGCACCACCTGTAACCAAAACAATTTTAGAAGTACTTTCACTTTGTAACTGAGCTGTCATTTATATTGAATTTTCAAAAATTCTTACTTTTTTTTTGGATGGACGCAAAGAAGTTTAATAATCATATTACATGGCA
TTACCACCATATACATATCCATATACATATCCATATCTAATCTTACTTATATGTTGTGGAAATGTAAAGAGCCCCATTATCTTAGCCTAAAAAAACCTTCTCTTTGGAACTTTCAGTAATACGCTTAACTGCTCATTGCTATA
TTGAAGTACGGATTAGAAGCCGCCGAGCGGGTGACAGCCCTCCGAAGGAAGACTCTCCTCCGTGCGTCCTCGTCTTCACCGGTCGCGTTCCTGAAACGCAGATGTGCCTCGCGCCGCACTGCTCCGAACAATAAAGATTCTAC
AATACTAGCTTTTATGGTTATGAAGAGGAAAAATTGGCAGTAACCTGGCCCCACAAACCTTCAAATGAACGAATCAAATTAACAACCATAGGATGATAATGCGATTAGTTTTTTAGCCTTATTTCTGGGGTAATTAATCAGC
GAAGCGATGATTTTTGATCTATTAACAGATATATAAATGCAAAAACTGCATAACCACTTTAACTAATACTTTCAACATT
TTCGGTTTGTATTACTTCTTATTCAAATGTAATAAAAGTATCAACAAAAAAATTGTTAATATACCTCTATACTTTAACGT
CAAGGAGAAAAAACTATAGAATTCTACTTCATACATTTTCAATTAAGATGCAGTTACTTCGCTGTTTTTCAATATTTTC
TGTTATTGCTTCAGTTTTAGCACAGGAACTGACAACTATATGCGAGCAAATCCCCTCACCAACTTTAGAATCGACGCCG
TACTCTTTGTCAACGACTACTATTTTGGCCAACGGGAAGGCAATGCAAGGAGTTTTTGAATATTACAAATCAGTAACGT
TTGTCAGTAATTGCGGTTCTCACCCCTCAACAACTAGCAAAGGCAGCCCCATAAACACACAGTATGTTTTTAAGGACAA
TAGCTCGACGATTGAAGGTAGATACCCATACGACGTTCCAGACTACGCTCTGCAGGCTAGTGGTGGTGGTGGTTCTGGT
GGTGGTGGTTCTGGTGGTGGTGGTTCTGCTAGCGTTTCTGATGTTCCGAGGGACCTGGAAGTTGTTGCTGCGACCCCCA
CCAGCCTACTGATCAGCTGGCCATGGATCTGCCCCGGGATGTACCATATGCCAATTTCCATTAATTACCGAACAGAAAT
TGACAAACCATCCCAGGGATCCGAACAAAAGCTTATTTCTGAAGAGGACTTGTAATAGCTCGAGATCTGATAACAACAG
TGTAGATGTAACAAAATCGACTTTGTTCCCACTGTACTTTTAGCTCGTACAAAATACAATATACTTTTCATTTCTCCGT
AAACAACATGTTTTCCCATGTAATATCCTTTTCTATTTTTCGTTCCGTTACCAACTTTACACATACTTTATATAGCTATT
CACTTCTATACACTAAAAAACTAAGACAATTTTAATTTTGCTGCCTGCCATATTTCAATTTGTTATAAATTCCTATAATTTATCCTATTAGTAGCTAAAAAAAGATGAATGTGAATCGAATCCTAAGAGAATTGAGCTCCAAT
TCGCCCTATAGTGAGTCGTATTACAATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCAC
CGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGCGCGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCG
CTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGG
CCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGC
CTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACAAAATATTAACGTTTACAATTTCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCAGGCAAGTGCACAAACAATA
CTTAAATAAATACTACTCAGTAATAACCTATTTCTTAGCATTTTTGACGAAATTTGCTATTTTGTTAGAGTCTTTTACACCATTTGTCTCCACACCTCCGCTTACATCAACACCAATAACGCCATTTAATCTAAGCGCATCAC
CAACATTTTCTGGCGTCAGTCCACCAGCTAACATAAAATGTAAGCTTTCGGGGCTCTCTTGCCTTCCAACCCAGTCAGAAATCGAGTTCCAATCCAAAAGTTCACCTGTCCCACCTGCTTCTGAATCAAACAAGGGAATAAAC
GAATGAGGTTTCTGTGAAGCTGCACTGAGTAGTATGTTGCAGTCTTTTGGAAATACGAGTCTTTTAATAACTGGCAAACCGAGGAACTCTTGGTATTCTTGCCACGACTCATCTCCATGCAGTTGGACGATATCAATGCCGTA
ATCATTGACCAGAGCCAAAACATCCTCCTTAGGTTGATTACGAAACACGCCAACCAAGTATTTCGGAGTGCCTGAACTATTTTTATATGCTTTTACAAGACTTGAAATTTTCCTTGCAATAACCGGGTCAATTGTTCTCTTTC
TATTGGGCACACATATAATACCCAGCAAGTCAGCATCGGAATCTAGAGCACATTCTGCGGCCTCTGTGCTCTGCAAGCCGCAAACTTTCACCAATGGACCAGAACTACCTGTGAAATTAATAACAGACATACTCCAAGCTGCC
TTTGTGTGCTTAATCACGTATACTCACGTGCTCAATAGTCACCAATGCCCTCCCTCTTGGCCCTCTCCTTTTCTTTTTTCGACCGAATTAATTCTTAATCGGCAAAAAAAGAAAAGCTCCGGATCAAGATTGTACGTAAGGTG
ACAAGCTATTTTTCAATAAAGAATATCTTCCACTACTGCCATCTGGCGTCATAACTGCAAAGTACACATATATTACGATGCTGTCTATTAAATGCTTCCTATATTATATATATAGTAATGTCGTTTATGGTGCACTCTCAGTA
CAATCTGCTCTGATGCCGCATAGTTAAGCCAGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCAGAGGTTT
TCACCGTCATCACCGAAACGCGCGA
```

pETh-Fn3wt

This construct is used for bacterial expression of Fn3 with a C-terminal His$_6$ tag.

Size: 5558 bp
Selectable marker: kanamycin

.... -- rbs -- TATA -- NdeI -- NheI -- Fn3 1-101 -- BamHI -- His6 -- Stop -- TTAACTAAACGA -- GATC....

```
TGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTT
CCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGT
TCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTA
TTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACAAAATATTAACGTTTACAATTTCAGGTGGCA
CTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAATTAATTCTTAGAAAAACTCATCGAGCATCAAATGAAACTGCAA
TTTATTCATATCAGGATTATCAATACCATATTTTTGAAAAAGCCGTTTCTGTAATGAAGGAGAAAACTCACCGAGGCAGTTCCATAGGATGGCAAGATCCTGGTATCGGTCTGCGATTC
CGACTCGTCCAACATCAATACAACCTATTAATTTCCCCTCGTCAAAAATAAGGTTATCAAGTGAGAAATCACCATGAGTGACGACTGAATCCGGTGAGAATGGCAAAAGTTTATGCATT
TCTTTCCAGACTTGTTCAACAGGCCAGCCATTACGCTCGTCATCAAAATCACTCGCATCAACCAAACCGTTATTCATTCGTGATTGCGCCTGAGCGAGACGAAATACGCGATCGCTGTT
AAAAGGACAATTACAAACAGGAATCGAATGCAACCGGCGCAGGAACACTGCCAGCGCATCAACAATATTTTCACCTGAATCAGGATATTCTTCTAATACCTGGAATGCTGTTTTCCCGG
GGATCGCAGTGGTGAGTAACCATGCATCATCAGGAGTACGGATAAAATGCTTGATGGTCGGAAGAGGCATAAATTCCGTCAGCCAGTTTAGTCTGACCATCTCATCTGTAACATCATTG
GCAACGCTACCTTTGCCATGTTTCAGAAACAACTCTGGCGCATCGGGCTTCCCATACAATCGATAGATTGTCGCACCTGATTGCCCGACATTATCGCGAGCCCATTTATACCCATATAA
ATCAGCATCCATGTTGGAATTTAATCGCGGCCTAGAGCAAGACGTTTCCCGTTGAATATGGCTCATAACACCCCTTGTATTACTGTTTATGTAAGCAGACAGTTTTATTGTTCATGACC
AAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACC
ACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAG
GCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAG
TTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCAC
GCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTC
GCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCAC
ATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGC
GGAAGAGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCATATATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATACACT
CCGCTATCGCTACGTGACTGGGTCATGGCTGCGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTC
CGGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGAGGCAGCTGCGGTAAAGCTCATCAGCGTGGTCGTGAAGCGATTCACAGATGTCTGCCTGTTCATCCGCGT
CCAGCTCGTTGAGTTTCTCCAGAAGCGTTAATGTCTGGCTTCTGATAAAGCGGGCCATGTTAAGGGCGGTTTTTTCCTGTTTGGTCACTGATGCCTCCGTGTAAGGGGGATTTCTGTTC
ATGGGGGTAATGATACCGATGAAACGAGAGAGGATGCTCACGATACGGGTTACTGATGATGAACATGCCCGGTTACTGGAACGTTGTGAGGGTAAACAACTGGCGGTATGGATGCGGCG
GGACCAGAGAAAAATCACTCAGGGTCAATGCCAGCGCTTCGTTAATACAGATGTAGGTGTTCCACAGGGTAGCCAGCAGCATCCTGCGATGCAGATCCGGAACATAATGGTGCAGGGCG
CTGACTTCCGCGTTTCCAGACTTTACGAAACACGGAAACCGAAGACCATTCATGTTGTTGCTCAGGTCGCAGACGTTTTGCAGCAGCAGTCGCTTCACGTTCGCTCGCGTATCGGTGAT
TCATTCTGCTAACCAGTAAGGCAACCCCGCCAGCCTAGCCGGGTCCTCAACGACAGGAGCACGATCATGCGCACCCGTGGGGCCGCCATGCCGGCGATAATGGCCTGCTTCTCGCCGAA
ACGTTTGGTGGCGGGACCAGTGACGAAGGCTTGAGCGAGGGCGTGCAAGATTCCGAATACCGCAAGCGACAGGCCGATCATCGTCGCGCTCCAGCGAAAGCGGTCCTCGCCGAAAATGA
CCCAGAGCGCTGCCGGCACCTGTCCTACGAGTTGCATGATAAAGAAGACAGTCATAAGTGCGGCGACGATAGTCATGCCCCGCGCCCACCGGAAGGAGCTGACTGGGTTGAAGGCTCTC
AAGGGCATCGGTCGAGATCCCGGTGCCTAATGAGTGAGCTAACTTACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGG
CCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCCAGGGTGGTTTTTCTTTTCACCAGTGAGACGGGCAACAGCTGATTGCCCTTCACCGCCTGGCCCTGAGAGAGTTGCAGCAAG
CGGTCCACGCTGGTTTGCCCCAGCAGGCGAAAATCCTGTTTGATGGTGGTTAACGGCGGGATATAACATGAGCTGTCTTCGGTATCGTCGTATCCCACTACCGAGATATCCGCACCAAC
GCGCAGCCCGGACTCGGTAATGGCGCGCATTGCGCCCAGCGCCATCTGATCGTTGGCAACCAGCATCGCAGTGGGAACGATGCCCTCATTCAGCATTTGCATGGTTTGTTGAAAACCGG
ACATGGCACTCCAGTCGCCTTCCCGTTCCGCTATCGGCTGAATTTGATTGCGAGTGAGATATTTATGCCAGCCAGCCAGACGCAGACGCGCCGAGACAGAACTTAATGGGCCCGCTAAC
AGCGCGATTTGCTGGTGACCCAATGCGACCAGATGCTCCACGCCCAGTCGCGTACCGTCTTCATGGGAGAAAATAATACTGTTGATGGGTGTCTGGTCAGAGACATCAAGAAATAACGC
CGGAACATTAGTGCAGGCAGCTTCCACAGCAATGGCATCCTGGTCATCCAGCGGATAGTTAATGATCAGCCCACTGACGCGTTGCGCGAGAAGATTGTGCACCGCCGCTTTACAGGCTT
CGACGCCGCTTCGTTCTACCATCGACACCACCACGCTGGCACCCAGTTGATCGGCGCGAGATTTAATCGCCGCGACAATTTGCGACGGCGCGTGCAGGGCCAGACTGGAGGTGGCAACG
CCAATCAGCAACGACTGTTTGCCCGCCAGTTGTTGTGCCACGCGGTTGGGAATGTAATTCAGCTCCGCCATCGCCGCTTCCACTTTTTCCCGCGTTTTCGCAGAAACGTGGCTGGCCTG
GTTCACCACGCGGGAAACGGTCTGATAAGAGACACCGGCATACTCTGCGACATCGTATAACGTTACTGGTTTCACATTCACCACCCTGAATTGACTCTCTTCCGGGCGCTATCATGCCA
TACCGCGAAAGGTTTTGCGCCATTCGATGGTGTCCGGGATCTCGACGCTCTCCCTTATGCGACTCCTGCATTAGGAAGCAGCCCAGTAGTAGGTTGAGGCCGTTGAGCACCGCCGCCGC
AAGGAATGGTGCATGCAAGGAGATGGCGCCCAACAGTCCCCCGGCCACGGGGCCTGCCACCATACCCACGCCGAAACAAGCGCTCATGAGCCCGAAGTGGCGAGCCCGATCTTCCCCAT
CGGTGATGTCGGCGATATAGGCGCCAGCAACCGCACCTGTGGCGCCGGTGATGCCGGCCACGATGCGTCCGGCGTAGAG
GATCGAGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAA
TAATTTTGTTTAACTTTAAGAAGGAGATATACATATGGCTAGCGTTTCTGATGTTCCGAGGGACCTGGAAGTTGTTGCT
GCGACCCCCACCAGCCTACTGATCAGCTGGGATGCTCCTGCTGTCACAGTGAGATATTACAGGATCACTTACGGAGAAA
CAGGAGGAAATAGCCCTGTCCAGGAGTTCACTGTGCCTGGGAGCAAGTCTACAGCTACCATCAGCGGCCTTAAACCTGG
AGTTGATTATACCATCACTGTGTATGCTGTCACTGGCCGTGGAGACAGCCCCGCAAGCAGCAAGCCAATTTCCATTAAT
TACCGAACAGAAATTGACAAACCATCCCAGGGATCCCACCATCACCATCATCACTGATTAACTAAACGAGATCCGGCTG
CTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAA
ACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAGGAACTATATCCGGAT
```

Translation from ATG in NdeI site onward:
M - AS - Fn3_{101} - GS - HHHHHH

MASVSDVPRDLEVVAATPTSLLISWDAPAVTVRYYRITYGETGGNSPVQEFTVPGSKST
ATISGLKPGVDYTITVYAVTGRGDSPASSKPISINYRTEIDKPSQGSHHHHHHZ

pEThK-Fn3wt

This construct is used for bacterial expression of Fn3 with a C-terminal His$_6$-KGSGK tag which provides two additional primary amines for chemical conjugation.

Size: 5573 bp
Selectable marker: kanamycin

```
.... -- rbs -- TATA -- NdeI -- NheI -- Fn3 1-101 -- BamHI -- His6  -- KGSGK -- Stop --
TTAACTAAACGA -- GATC....
```

```
TGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTT
CCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGT
TCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTA
TTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACAAAATATTAACGTTTACAATTTCAGGTGGCA
CTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAATTAATTCTTAGAAAAACTCATCGAGCATCAAATGAAACTGCAA
TTTATTCATATCAGGATTATCAATACCATATTTTTGAAAAAGCCGTTTCTGTAATGAAGGAGAAAACTCACCGAGGCAGTTCCATAGGATGGCAAGATCCTGGTATCGGTCTGCGATTC
CGACTCGTCCAACATCAATACAACCTATTAATTTCCCCTCGTCAAAAATAAGGTTATCAAGTGAGAAATCACCATGAGTGACGACTGAATCCGGTGAGAATGGCAAAAGTTTATGCATT
TCTTTCCAGACTTGTTCAACAGGCCAGCCATTACGCTCGTCATCAAAATCACTCGCATCAACCAAACCGTTATTCATTCGTGATTGCGCCTGAGCGAGACGAAATACGCGATCGCTGTT
AAAAGGACAATTACAAACAGGAATCGAATGCAACCGGCGCAGGAACACTGCCAGCGCATCAACAATATTTTCACCTGAATCAGGATATTCTTCTAATACCTGGAATGCTGTTTTCCCGG
GGATCGCAGTGGTGAGTAACCATGCATCATCAGGAGTACGGATAAAATGCTTGATGGTCGGAAGAGGCATAAATTCCGTCAGCCAGTTTAGTCTGACCATCTCATCTGTAACATCATTG
GCAACGCTACCTTTGCCATGTTTCAGAAACAACTCTGGCGCATCGGGCTTCCCATACAATCGATAGATTGTCGCACCTGATTGCCCGACATTATCGCGAGCCCATTTATACCCATATAA
ATCAGCATCCATGTTGGAATTTAATCGCGGCCTAGAGCAAGACGTTTCCCGTTGAATATGGCTCATAACACCCCTTGTATTACTGTTTATGTAAGCAGACAGTTTTATTGTTCATGACC
AAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACC
ACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAG
GCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAG
TTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCAC
GCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTC
GCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCAC
ATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGC
GGAAGAGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCATATATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATACACT
CCGCTATCGCTACGTGACTGGGTCATGGCTGCGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTC
CGGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGAGGCAGCTGCGGTAAAGCTCATCAGCGTGGTCGTGAAGCGATTCACAGATGTCTGCCTGTTCATCCGCGT
CCAGCTCGTTGAGTTTCTCCAGAAGCGTTAATGTCTGGCTTCTGATAAAGCGGGCCATGTTAAGGGCGGTTTTTTCCTGTTTGGTCACTGATGCCTCCGTGTAAGGGGGATTTCTGTTC
ATGGGGGTAATGATACCGATGAAACGAGAGAGGATGCTCACGATACGGGTTACTGATGATGAACATGCCCGGTTACTGGAACGTTGTGAGGGTAAACAACTGGCGGTATGGATGCGGCG
GGACCAGAGAAAAATCACTCAGGGTCAATGCCAGCGCTTCGTTAATACAGATGTAGGTGTTCCACAGGGTAGCCAGCAGCATCCTGCGATGCAGATCCGGAACATAATGGTGCAGGGCG
CTGACTTCCGCGTTTCCAGACTTTACGAAACACGGAAACCGAAGACCATTCATGTTGTTGCTCAGGTCGCAGACGTTTTGCAGCAGCAGTCGCTTCACGTTCGCTCGCGTATCGGTGAT
TCATTCTGCTAACCAGTAAGGCAACCCCGCCAGCCTAGCCGGGTCCTCAACGACAGGAGCACGATCATGCGCACCCGTGGGGCCGCCATGCCGGCGATAATGGCCTGCTTCTCGCCGAA
ACGTTTGGTGGCGGGACCAGTGACGAAGGCTTGAGCGAGGGCGTGCAAGATTCCGAATACCGCAAGCGACAGGCCGATCATCGTCGCGCTCCAGCGAAAGCGGTCCTCGCCGAAAATGA
CCCAGAGCGCTGCCGGCACCTGTCCTACGAGTTGCATGATAAAGAAGACAGTCATAAGTGCGGCGACGATAGTCATGCCCCGCGCCCACCGGAAGGAGCTGACTGGGTTGAAGGCTCTC
AAGGGCATCGGTCGAGATCCCGGTGCCTAATGAGTGAGCTAACTTACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGG
CCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCCAGGGTGGTTTTTCTTTTCACCAGTGAGACGGGCAACAGCTGATTGCCCTTCACCGCCTGGCCCTGAGAGAGTTGCAGCAAG
CGGTCCACGCTGGTTTGCCCCAGCAGGCGAAAATCCTGTTTGATGGTGGTTAACGGCGGGATATAACATGAGCTGTCTTCGGTATCGTCGTATCCCACTACCGAGATATCCGCACCAAC
GCGCAGCCCGGACTCGGTAATGGCGCGCATTGCGCCCAGCGCCATCTGATCGTTGGCAACCAGCATCGCAGTGGGAACGATGCCCTCATTCAGCATTTGCATGGTTTGTTGAAAACCGG
ACATGGCACTCCAGTCGCCTTCCCGTTCCGCTATCGGCTGAATTTGATTGCGAGTGAGATATTTATGCCAGCCAGCCAGACGCAGACGCGCCGAGACAGAACTTAATGGGCCCGCTAAC
AGCGCGATTTGCTGGTGACCCAATGCGACCAGATGCTCCACGCCCAGTCGCGTACCGTCTTCATGGGAGAAAATAATACTGTTGATGGGTGTCTGGTCAGAGACATCAAGAAATAACGC
CGGAACATTAGTGCAGGCAGCTTCCACAGCAATGGCATCCTGGTCATCCAGCGGATAGTTAATGATCAGCCCACTGACGCGTTGCGCGAGAAGATTGTGCACCGCCGCTTTACAGGCTT
CGACGCCGCTTCGTTCTACCATCGACACCACCACGCTGGCACCCAGTTGATCGGCGCGAGATTTAATCGCCGCGACAATTTGCGACGGCGCGTGCAGGGCCAGACTGGAGGTGGCAACG
CCAATCAGCAACGACTGTTTGCCCGCCAGTTGTTGTGCCACGCGGTTGGGAATGTAATTCAGCTCCGCCATCGCCGCTTCCACTTTTTCCCGCGTTTTCGCAGAAACGTGGCTGGCCTG
GTTCACCACGCGGGAAACGGTCTGATAAGAGACACCGGCATACTCTGCGACATCGTATAACGTTACTGGTTTCACATTCACCACCCTGAATTGACTCTCTTCCGGGCGCTATCATGCCA
TACCGCGAAAGGTTTTGCGCCATTCGATGGTGTCCGGGATCTCGACGCTCTCCCTTATGCGACTCCTGCATTAGGAAGCAGCCCAGTAGTAGGTTGAGGCCGTTGAGCACCGCCGCCGC
AAGGAATGGTGCATGCAAGGAGATGGCGCCCAACAGTCCCCCGGCCACGGGGCCTGCCACCATACCCACGCCGAAACAAGCGCTCATGAGCCCGAAGTGGCGAGCCCGATCTTCCCCAT
CGGTGATGTCGGCGATATAGGCGCCAGCAACCGCACCTGTGGCGCCGGTGATGCCGGCCACGATGCGTCCGGCGTAGAG
GATCGAGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAA
TAATTTTGTTTAACTTTAAGAAGGAGATATACATATGGCTAGCGTTTCTGATGTTCCGAGGGACCTGGAAGTTGTTGCT
GCGACCCCCACCAGCCTACTGATCAGCTGGGATGCTCCTGCTGTCACAGTGAGATATTACAGGATCACTTACGGAGAAA
CAGGAGGAAATAGCCCTGTCCAGGAGTTCACTGTGCCTGGGAGCAAGTCTACAGCTACCATCAGCGGCCTTAAACCTGG
AGTTGATTATACCATCACTGTGTATGCTGTCACTGGCCGTGGAGACAGCCCCGCAAGCAGCAAGCCAATTTCCATTAAT
TACCGAACAGAAATTGACAAACCATCCCAGGGATCCCACCATCACCATCATCACAAGGGTTCTGGCAAATGATTAACTA
AACGAGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACC
CCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAGGAACTATATCCGGAT
```

Translation from ATG in NdeI site onward:

M - AS - Fn3$_{101}$ - GS - HHHHHH - KGSGK

```
MASVSDVPRDLEVVAATPTSLLISWDAPAVTVRYYRITYGETGGNSPVQEFTVPGSKST
ATISGLKPGVDYTITVYAVTGRGDSPASSKPISINYRTEIDKPSQGSHHHHHHKGSGKZ
```

pETh-Fn3-Fn3

This construct is used for bacterial expression of Fn3-Fn3 bivalent with a C-terminal His$_6$ tag.

.... - rbs - TATA - NdeI - NheI - Fn3$_{101}$ - BamHI - linker -
KpnI - Fn3$_{101}$ - SacI - spacer - His6 - Stop - TTAACTAAACGA -
- GATC....

```
TGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTT
CCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGT
TCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTA
TTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACAAAATATTAACGTTTACAATTTCAGGTGGCA
CTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAATTAATTCTTAGAAAAACTCATCGAGCATCAAATGAAACTGCAA
TTTATTCATATCAGGATTATCAATACCATATTTTTGAAAAAGCCGTTTCTGTAATGAAGGAGAAAACTCACCGAGGCAGTTCCATAGGATGGCAAGATCCTGGTATCGGTCTGCGATTC
CGACTCGTCCAACATCAATACAACCTATTAATTTCCCCTCGTCAAAAATAAGGTTATCAAGTGAGAAATCACCATGAGTGACGACTGAATCCGGTGAGAATGGCAAAAGTTTATGCATT
TCTTTCCAGACTTGTTCAACAGGCCAGCCATTACGCTCGTCATCAAAATCACTCGCATCAACCAAACCGTTATTCATTCGTGATTGCGCCTGAGCGAGACGAAATACGCGATCGCTGTT
AAAAGGACAATTACAAACAGGAATCGAATGCAACCGGCGCAGGAACACTGCCAGCGCATCAACAATATTTTCACCTGAATCAGGATATTCTTCTAATACCTGGAATGCTGTTTTCCCGG
GGATCGCAGTGGTGAGTAACCATGCATCATCAGGAGTACGGATAAAATGCTTGATGGTCGGAAGAGGCATAAATTCCGTCAGCCAGTTTAGTCTGACCATCTCATCTGTAACATCATTG
GCAACGCTACCTTTGCCATGTTTCAGAAACAACTCTGGCGCATCGGGCTTCCCATACAATCGATAGATTGTCGCACCTGATTGCCCGACATTATCGCGAGCCCATTTATACCCATATAA
ATCAGCATCCATGTTGGAATTTAATCGCGGCCTAGAGCAAGACGTTTCCCGTTGAATATGGCTCATAACACCCCTTGTATTACTGTTTATGTAAGCAGACAGTTTTATTGTTCATGACC
AAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACC
ACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAG
GCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAG
TTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCAC
GCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTC
GCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCAC
ATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGC
GGAAGAGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCATATATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATACACT
CCGCTATCGCTACGTGACTGGGTCATGGCTGCGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTC
CGGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGAGGCAGCTGCGGTAAAGCTCATCAGCGTGGTCGTGAAGCGATTCACAGATGTCTGCCTGTTCATCCGCGT
CCAGCTCGTTGAGTTTCTCCAGAAGCGTTAATGTCTGGCTTCTGATAAAGCGGGCCATGTTAAGGGCGGTTTTTTCCTGTTTGGTCACTGATGCCTCCGTGTAAGGGGGATTTCTGTTC
ATGGGGGTAATGATACCGATGAAACGAGAGAGGATGCTCACGATACGGGTTACTGATGATGAACATGCCCGGTTACTGGAACGTTGTGAGGGTAAACAACTGGCGGTATGGATGCGGCG
GGACCAGAGAAAAATCACTCAGGGTCAATGCCAGCGCTTCGTTAATACAGATGTAGGTGTTCCACAGGGTAGCCAGCAGCATCCTGCGATGCAGATCCGGAACATAATGGTGCAGGGCG
CTGACTTCCGCGTTTCCAGACTTTACGAAACACGGAAACCGAAGACCATTCATGTTGTTGCTCAGGTCGCAGACGTTTTGCAGCAGCAGTCGCTTCACGTTCGCTCGCGTATCGGTGAT
TCATTCTGCTAACCAGTAAGGCAACCCCGCCAGCCTAGCCGGGTCCTCAACGACAGGAGCACGATCATGCGCACCCGTGGGGCCGCCATGCCGGCGATAATGGCCTGCTTCTCGCCGAA
ACGTTTGGTGGCGGGACCAGTGACGAAGGCTTGAGCGAGGGCGTGCAAGATTCCGAATACCGCAAGCGACAGGCCGATCATCGTCGCGCTCCAGCGAAAGCGGTCCTCGCCGAAAATGA
CCCAGAGCGCTGCCGGCACCTGTCCTACGAGTTGCATGATAAAGAAGACAGTCATAAGTGCGGCGACGATAGTCATGCCCCGCGCCCACCGGAAGGAGCTGACTGGGTTGAAGGCTCTC
AAGGGCATCGGTCGAGATCCCGGTGCCTAATGAGTGAGCTAACTTACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGG
CCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCCAGGGTGGTTTTTCTTTTCACCAGTGAGACGGGCAACAGCTGATTGCCCTTCACCGCCTGGCCCTGAGAGAGTTGCAGCAAG
CGGTCCACGCTGGTTTGCCCCAGCAGGCGAAAATCCTGTTTGATGGTGGTTAACGGCGGGATATAACATGAGCTGTCTTCGGTATCGTCGTATCCCACTACCGAGATATCCGCACCAAC
GCGCAGCCCGGACTCGGTAATGGCGCGCATTGCGCCCAGCGCCATCTGATCGTTGGCAACCAGCATCGCAGTGGGAACGATGCCCTCATTCAGCATTTGCATGGTTTGTTGAAAACCGG
ACATGGCACTCCAGTCGCCTTCCCGTTCCGCTATCGGCTGAATTTGATTGCGAGTGAGATATTTATGCCAGCCAGCCAGACGCAGACGCGCCGAGACAGAACTTAATGGGCCCGCTAAC
AGCGCGATTTGCTGGTGACCCAATGCGACCAGATGCTCCACGCCCAGTCGCGTACCGTCTTCATGGGAGAAAATAATACTGTTGATGGGTGTCTGGTCAGAGACATCAAGAAATAACGC
CGGAACATTAGTGCAGGCAGCTTCCACAGCAATGGCATCCTGGTCATCCAGCGGATAGTTAATGATCAGCCCACTGACGCGTTGCGCGAGAAGATTGTGCACCGCCGCTTTACAGGCTT
CGACGCCGCTTCGTTCTACCATCGACACCACCACGCTGGCACCCAGTTGATCGGCGCGAGATTTAATCGCCGCGACAATTTGCGACGGCGCGTGCAGGGCCAGACTGGAGGTGGCAACG
CCAATCAGCAACGACTGTTTGCCCGCCAGTTGTTGTGCCACGCGGTTGGGAATGTAATTCAGCTCCGCCATCGCCGCTTCCACTTTTTCCCGCGTTTTCGCAGAAACGTGGCTGGCCTG
GTTCACCACGCGGGAAACGGTCTGATAAGAGACACCGGCATACTCTGCGACATCGTATAACGTTACTGGTTTCACATTCACCACCCTGAATTGACTCTCTTCCGGGCGCTATCATGCCA
TACCGCGAAAGGTTTTGCGCCATTCGATGGTGTCCGGGATCTCGACGCTCTCCCTTATGCGACTCCTGCATTAGGAAGCAGCCCAGTAGTAGGTTGAGGCCGTTGAGCACCGCCGCCGC
AAGGAATGGTGCATGCAAGGAGATGGCGCCCAACAGTCCCCCGGCCACGGGGCCTGCCACCATACCCACGCCGAAACAAGCGCTCATGAGCCCGAAGTGGCGAGCCCGATCTTCCCCAT
CGGTGATGTCGGCGATATAGGCGCCAGCAACCGCACCTGTGGCGCCGGTGATGCCGGCCACGATGCGTCCGGCGTAGAG
GATCGAGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAA
TAATTTTGTTTAACTTTAAGAAGGAGATATACATATGGCTAGCGTTTCTGATGTTCCGAGGGACCTGGAAGTTGTTGCT
GCGACCCCCACCAGCCTACTGATCAGCTGGCTTCACCATCGCTCTGACGTGCGCTCTTACAGGATCACTTACGGAGAAA
CAGGAGGAAATAGCCCTGTCCAGAAGTTCACTGTGCCTGGGTCGCGCTCCCTGGCTACCATCAGCGGCCTTAAACCTGG
AGTTGATTATACCATCACTGTGTATGCTGTCACTTGGGGGTCTTACTGTTGCTCTAATCCAATTTCCATTAATTACCGA
ACAGAAATTGACAAACCATCCCAGGGATCCGGAGGCGGTTCAGGCGGAGGTAAAGGTGGCGGAGGTACCGTTTCTGATG
TTCCGAGGGACCTGGAAGTTGTTGCTGCGACCCCCACCAGCCTACTGATCAGCTGGTATCATCCTTTCTATTATGTCGC
GCATTCTTACAGGATCACTTACGGAGAAACAGGAGGAAATAGCCCTGTCCAGGAGTTCACTGTGCCTCGTTCGCCCTGG
TTTGCTACCATCAGCGGCCTTAAACCTGGAGTTGATTATACCATCACTGTGTATGCTGTCACTGATAGTAACGGTTCTC
ATCCAATTTCCATTAATTACCGAACAGAAATTGACAAACCATCCCAGGAGCTCAGATCCCACCATCACCATCATCACTG
ATTAACTAAACGAGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTA
GCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAGGAACTATATCCGGAT
```

Translation from ATG in NdeI site onward:

M - AS - Fn3$_{101}$ - linker - Fn3$_{101}$ - ELRS - HHHHHH

MASVSDVPRDLEVVAATPTSLLISWLHHRSDVRSYRITYGETGGNSPVQKFTVPGSRSL
ATISGLKPGVDYTITVYAVTWGSYCCSNPISINYRTEIDKPSQGSGGGSGGGKGGGGTV
SDVPRDLEVVAATPTSLLISWYHPFYYVAHSYRITYGETGGNSPVQEFTVPRSPWFATI
SGLKPGVDYTITVYAVTDSNGSHPISINYRTEIDKPSQELRSHHHHHH*

pCT-EGFR(404SG)

This construct is used for yeast surface display of EGFR ectodomain mutant 404SG.

Size: 8129 bp
Selectable marker: ampicillin

Aga2--Spacer--FactorXa--HA--Linker--NheI--404SG--BamHI--
Myc--Stop2--XhoI--Terminator....

ACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGATAATAATGGTTTCTTAGGACGGATCGCTTGCCTGTAACTTACACGCGCCTCGTATCTTTTAATGATGGAATAATTTGGGAATTTACTCTGTGTTTA
TTTATTTTTATGTTTTGTATTTGGATTTTAGAAAGTAAATAAAGAAGGTAGAAGAGTTACGGAATGAAGAAAAAAAAATAAACAAAGGTTTAAAAAATTTCAACAAAAAGCGTACTTTACATATATATTTATTAGACAAGAAA
AGCAGATTAAATAGATATACATTCGATTAACGATAAGTAAAATGTAAAATCACAGGATTTTCGTGTGTGGTCTTCTACACAGACAAGATGAAACAATTCGGCATTAATACCTGAGAGCAGGAAGAGCAAGATAAAAGGTAGTA
TTTGTTGGCGATCCCCCTAGAGTCTTTTACATCTTCGGAAAACAAAAACTATTTTTTCTTTAATTTCTTTTTTTACTTTCTATTTTTAATTTATATATTTATATTAAAAAATTTAAATTATAATTATTTTTATAGCACGTGAT
GAAAAGGACCCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTAT
GAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATC
TCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCT
CAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAA
GGAGCTAACCGCTTTTTTTCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTAT
TAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGG
TCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGCAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTA
ACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAG
ACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACT
GGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTC
GTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCATTGAGAAA
GCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGGAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAG
CGTCGATTTTTGTGATGCTCGTCAGGGGGGCCGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCG
TATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGC
ACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTACCTCACTCATTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCCTATGTTGTGTGGAATTGTGAGCGGATAACAATTTCAC
ACAGGAAACAGCTATGACCATGATTACGCCAAGCTCGGAATTAACCCTCACTAAAGGGAACAAAAGCTGGGTACCCGACAGGTTATCAGCAACAACACAGTCATATCCATTCTCAATTAGCTCTACCACAGTGTGTGAACCAA
TGTATCCAGCACCACCTGTAACCAAAACAATTTTAGAAGTACTTTCACTTTGTAACTGAGCTGTCATTTATATTGAATTTTCAAAAATTCTTACTTTTTTTTTGGATGGACGCAAAGAAGTTTAATAATCATATTACATGGCA
TTACCACCATATACATATCCATATACATATCCATATCTAATCTTACTTATATGTTGTGGAAATGTAAAGAGCCCCATTATCTTAGCCTAAAAAAACCTTCTCTTTGGAACTTTCAGTAATACGCTTAACTGCTCATTGCTATA
TTGAAGTACGGATTAGAAGCCGCCGAGCGGGTGACAGCCCTCCGAAGGAAGACTCTCCTCCGTGCGTCCTCGTCTTCACCGGTCGCGTTCCTGAAACGCAGATGTGCCTCGCGCCGCACTGCTCCGAACAATAAAGATTCTAC
AATACTAGCTTTTATGGTTATGAAGAGGAAAAATTGGCAGTAACCTGGCCCCACAAACCTTCAAATGAACGAATCAAATTAACAACCATAGGATGATAATGCGATTAGTTTTTTAGCCTTATTTCTGGGGTAATTAATCAGC
GAAGCGATGATTTTTGATCTATTAACAGATATATAAATGCAAAAACTGCATAACCACTTTAACTAATACTTTCAACATTTTCGGTTTGTATTACTTCTTATT
CAAATGTAATAAAAGTATCAACAAAAAATTGTTAATATACCTCTATACTTTAACGTCAAGGAGAAAAAACTATAGAATTCTACTTCATACATTTTCAATTAA
GATGCAGTTACTTCGCTGTTTTTCAATATTTTCTGTTATTGCTTCAGTTTTAGCACAGGAACTGACAACTATATGCGAGCAAATCCCCTCACCAACTTTAGA
ATCGACGCCGTACTCTTTGTCAACGACTACTATTTTGGCCAACGGGAAGGCAATGCAAGGAGTTTTTGAATATTACAAATCAGTAACGTTTGTCAGTAATTG
CGGTTCTCACCCCTCAACAACTAGCAAAGGCAGCCCCATAAACACACAGTATGTTTTTAAGGACAATAGCTCGACGATTGAAGGTAGATACCCATACGACGT
TCCAGACTACGCTCTGCAGGCTAGTGGTGGTGGTGGTTCTGGTGGTGGTGGTTCTGGTGGTGGTGGTTCTGCTAGCCTGGAGGAAAAGAAAGTTTGCCAAGG
CACGAGTAACAAGCTCACGCAGTTGGGCACTTTTGAAGATCATTTTCTCAGCCTCCAGAGGATGTTCAATAACTGTGAGGTGGTCCTTGGGAATTTGGAAAT
TACCTATGTGCAGAGGAATTATGATCTTTCCTTCTTAAAGACCATCCAGGAGGTGACTGGTTATGTCCTCATTGCCCACAACACAGTGGAGCGAATTCCTTT
GGAAAACCTGCAGATCATCAGAGGAAATATGTACTACGAAAATTCCTATGCCTTAGCAGTCTTATCTAACTATGATGCAAATAAAACCGGACTGAAGGAGCT
GCCCATGAGAAATTTACAGGAAATCCTGCATGGCGCCGTGCGGTTCAGCAACAACCCTGCCCTGTGCAACGTGGAGAGCATCCAGTGGCGGGACATAGTCAG
CAGTGACTTTCTCAGCAACATGTCGATGGACTTCCAGAACCACCTGGGCAGCTGCCAAAAGTGTGATCCAAGCTGTCCCAATGGGAGCTGCTGGGGTGCAGG
AGAGGAGAACTGCCAGAAACTGACCAAAATCATCTGTGCCCAGCAGTGCTCCGGGCGCTGCCGTGGCAAGTCCCCCAGTGACTGCTGCCACAACCAGTGTGC
TGCAGGCTGCACAGGCCCCCGGGAGAGCGACTGCCTGGTCTGCCGCAAATTCCGAGACGAAGCCACGTGCAAGGACACCTGCCCCCCACTCATGCTCTACAA
CCCCACCACGTACCAGATGGATGTGAACCCCGAGGGCAAATACAGCTTTGGTGCCACCTGCGTGAAGAAGTGTCCCCGTAATTATGTGGTGACAGATCACGG
CTCGTGCGTCCGAGCCTGTGGGGCCGACAGCTATGAGATGGAGGAAGACGGCGTCCGCAAGTGTAAGAAGTGCGAAGGGCCTTGCCGCAAAGTGTGTAACGG
AATAGGTATTGGTGAATTTAAAGACTCACTCTCCATAAATGCTACGAATATTAAACACTTCAAAAACTGCACCTCCATCAGTGGCGATCTCCACATCCTGCC
GGTGGCATTTAGGGGTGACTCCTTCACACATACTCCTCCTCTGGACCCACAGGAACTGGATATTCTGAAAACCGTAAAGGAAATCACAGGGTCTTTGCTGAT
TCAGGCTTGGCCTGAAAACAGGACGGACCTCCATGCCTTTGAGAACCTAGAAATCATACGCGGCAGGACCAAGCAACATGGTCAGTTTTCTCTTGCAGTCGT
CGGCCTGAACATAACATCCTTGGGATTACGCTCCCTCAAGGAGATAAGTGATGGAGATGTGATAATTTCAGGAAACAAAAATTTGTGCTATGCAAATACAAT
AAACTGGAAAAAACTGTTTGGGACCTCCGGTCAGAAAACCAAAATTATAAGCAACAGAGGTGAAAACAGCTGCAAGGCCACAGGCCAGGTCTGCCATGCCTT
GTGCTCCCCCGAGGGCTGCTGGGGCCCGGAGCCCAGGGACTGCGTCTCTTGCCGGAATGTCAGCCGAGGCAGGGAATGCGTGGACAAGTGCAACCTTCTGGA
GGGTGAGCCAAGGGAGTTTGTGGAGAACTCTGAGTGCATACAGTGCCACCCAGAGTGCCTGCCTCAGGCCATGAACATCACCTGCACAGGACGGGGACCAGA
CAACTGTATCCAGTGTGCCCACTACATTGACGGCCCCCACTGCGTCAAGACCTGCCCGGCAGGAGTCATGGGAGAAAACAACACCCTGGTCTGGAAGTACGC
AGACGCCGGCCATGTGTGCCACCTGTGCCATCCAAACTGCACCTACGGATGCACTGGGCCAGGTCTTGAAGGCTGTCCAACGAATGGGCCTAAGATCCCGTC
CGGATCCGAACAAAAGCTTATTTCTGAAGAGGACTTGTAATAGCTCGAGATCTGATAACAACAGTGTAGATGTAACAAAATCGACTTTGTTCCCACTGTACT
TTTAGCTCGTACAAAATACAATATACTTTTCATTTCTCCGTAAACAACATGTTTTCCCATGTAATATCCTTTTCTATTTTTCGTTCCGTTACCAACTTTACACATACTTTATATAGCTAT
TCACTTCTATACACTAAAAAACTAAGACAATTTTAATTTTGCTGCCTGCCATATTTCAATTTGTTATAAATTCCTATAATTTATCCTATTAGTAGCTAAAAAAAGATGAATGTGAATCGAATCCTAAGAGAATTGAGCTCCAA
TTCGCCCTATAGTGAGTCGTATTACAATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCCTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCA
CCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGCGCGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTC
GCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGG
GCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGG
CCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACAAAATATTAACGTTTACAATTTCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCAGGCAAGTGCACAAACAAT
ACTTAAATAAATACTACTCAGTAATAACCTATTTCTTAGCATTTTTGACGAAATTTGCTATTTTGTTAGAGTCTTTTACACCATTTGTCTCCACACCTCCGCTTACATCAACACCAATAACGCCATTTAATCTAAGCGCATCA
CCAACATTTTCTGGCGTCAGTCCACCAGCTAACATAAAATGTAAGCTTTCGGGGCTCTCTTGCCTTCCAACCCAGTCAGAAATCGAGTTCCAATCCAAAAGTTCACCTGTCCCACCTGCTTCTGAATCAAACAAGGGAATAAA
CGAATGAGGTTTCTGTGAAGCTGCACTGAGTAGTATGTTGCAGTCTTTTGGAAATACGAGTCTTTTAATAACTGGCAAACCGAGGAACTCTTGGTATTCTTGCCACGACTCATCTCCATGCAGTTGGACGATATCAATGCCGT
AATCATTGACCAGAGCCAAAACATCCTCCTTAGGTTGATTACGAAACACGCCAACCAAGTATTTCGGAGTGCCTGAACTATTTTTATATGCTTTTACAAGACTTGAAATTTTCCTTGCAATAACCGGGTCAATTGTTCTCTTT
CTATTGGGCACACATATAATACCCAGCAAGTCAGCATCGGAATCTAGAGCACATTCTGCGGCCTCTGTGCTCTGCAAGCCGCAAACTTTCACCAATGGACCAGAACTACCTGTGAAATTAATAACAGACATACTCCAAGCTGC
CTTTGTGTGCTTAATCACGTATACTCACGTGCTCAATAGTCACCAATGCCCTCCCTCTTGGCCCTCTCCTTTTCTTTTTTCGACCGAATTAATTCTTAATCGGCAAAAAAAGAAAAGCTCCGGATCAAGATTGTACGTAAGGT
GACAAGCTATTTTTCAATAAAGAATATCTTCCACTACTGCCATCTGGCGTCATAACTGCAAAGTACACATATATTACGATGCTGTCTATTAAATGCTTCCTATATTATATATATAGTAATGTCGTTTATGGTGCACTCTCAGT
ACAATCTGCTCTGATGCCGCATAGTTAAGCCAGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCAGAGGTT
TTCACCGTCATCACCGAAACGCGCGA

Translation from Aga2p onward:
Aga2p - KDNSST - Xa - HA - $(G_4S)_3$ - AS - 404SG - GS - c-myc

MQLLRCFSIFSVIASVLAQELTTICEQIPSPTLESTPYSLSTTTILANGKAMQGVFEYYKSVTFVSNCGSHPSTTSKGS
PINTQYVFKDNSSTIEGRYPYDVPDYALQASGGGGSGGGGSGGGGSASLEEKKVCQGTSNKLTQLGTFEDHFLSLQRMF
NNCEVVLGNLEITYVQRNYDLSFLKTIQEVTGYVLIAHNTVERIPLENLQIIRGNMYYENSYALAVLSNYDANKTGLKE
LPMRNLQEILHGAVRFSNNPALCNVESIQWRDIVSSDFLSNMSMDFQNHLGSCQKCDPSCPNGSCWGAGEENCQKLTKI
ICAQQCSGRCRGKSPSDCCHNQCAAGCTGPRESDCLVCRKFRDEATCKDTCPPLMLYNPTTYQMDVNPEGKYSFGATCV
KKCPRNYVVTDHGSCVRACGADSYEMEEDGVRKCKKCEGPCRKVCNGIGIGEFKDSLSINATNIKHFKNCTSISGDLHI
LPVAFRGDSFTHTPPLDPQELDILKTVKEITGSLLIQAWPENRTDLHAFENLEIIRGRTKQHGQFSLAVVGLNITSLGL
RSLKEISDGDVIISGNKNLCYANTINWKKLFGTSGQKTKIISNRGENSCKATGQVCHALCSPEGCWGPEPRDCVSCRNV
SRGRECVDKCNLLEGEPREFVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCPAGVMGENNTLVW
KYADAGHVCHLCHPNCTYGCTGPGLEGCPTNGPKIPSGSEQKLISEEDLZZ

Engineered Binders

These constructs are Fn3 domains engineered for binding to the indicated target. Sequence data is provided from NheI to BamHI in both nucleotide and amino acid formats.

Lysozyme

L0.7.1 (~~10 μM)
GCTAGCGTTTCTGATGTTCCGAGGGACCTGGAAGTTGTTGCTGCGACCCCCACCAGCCTACTGATCAGCTG
GCGCGACTGCCCCTGGGCTACCTATTACAGGATCACTTACGGAGAAACAGGAGGAAATAGCCCTGTCCAGG
AGTTCACTGTGCCTTGGACCCCCGTTTGTTTTGCTACCATCAGCGGCCTTAAACCTGGAGTTGATTATACC
ATCACTGTGTATGCTGTCACTTCTAGTCAGCGGGGCTGCATGCCAATTTCCATTAATTACCGAACAGAAAT
TGACAAACCATCCCAGGGATCC

ASVSDVPRDLEVVAATPTSLLISWRDCPWATYYRITYGETGGNSPVQEFTVPWTPVCFATISGLKPGVDYT
ITVYAVTSSQRGCMPISINYRTEIDKPSQGS

L1.5.1 (~1 μM)
GCTAGCGTTTCTGATGTTCCGAGGGACCTGGAAGTTGTTGCTGCGACCCCCACCAGCCTACTGATCAGCTG
GAGTTTGGACAACCAGGCCAATTATTACAGGATCACTTACGGAGAAACAGGAGGAAATAGCCCTGTCCAGG
AGTTCACTGTGCCTGGTCAGTCGGACGCTACCATCAGCGGCCTTAAACCTGGAGTTGATTATACCATCACT
GTGTATGCTGTCACTCGGTGTGAGCCGTCCCGCAATTCGGCCGTTCCAATTTCCATTAATTACCGAACAGA
AATTGACAAACCATCCCAGGGATCC

ASVSDVPRDLEVVAATPTSLLISWSLDNQANYYRITYGETGGNSPVQEFTVPGQSDATISGLKPGVDYTIT
VYAVTRCEPSRNSAVPISINYRTEIDKPSQGS

L3.3.1 (7.6 ± 1.1 nM)
GCTAGCGTTTCTGATGTTCCGAGGGACCTGGAAGTTGTTGCTGCGACCCCCACCAGCCTACTGATCAGCTG
GAGTTTGGACAACCAGGCCAATTATTACAGGATCACTTACGGAGAAACAGGAGGAAATAGCTCTGTCCAGG
AGTTCACTATGCCTGGGGTTACCAATGCTACCATCAGCGGCCTTAAACCTGGAGTTGATTATACCATCACT
GTGTATGCTGTCACTCGCGTGGGGCGGATGCTTGACACGCCGGGCCCAATTTCCATTAATTACCGAACAGA
AATTGACAAACCATCCCAGGGATCC

ASVSDVPRDLEVVAATPTSLLISWSLDNQANYYRITYGETGGNSSVQEFTMPGVTNATISGLKPGVDYTIT
VYAVTRVGRMLDTPGPISINYRTEIDKPSQGS

L5.3.4 (360 pM)
GCTAGCGTTTCTGATGTCCCGAGGGGCCTGGAAGTTGTTGCTGCGACCCCCACCAGCCTACTGATCAGCTG
GAGTTTGGACAACCAGGCCAAGTATTACAGGATCACTTACGGGGAAACAGGGGGAAATAGCCTTGTCCAGG
AGTTCACTATGCCTGGGGTTACCAATGCTACCATCAGCGGCCTTAAACCTGGAGTTGATTATACCATCACT
GTGTATGCTGTCACTCGCGTGGGGCGGATGCTTGACACGCCGGGCCCAATTTCCACTAATTACCGAACAGA
GATTGACAAACCATCCCAGGGATCC

ASVSDVPRGLEVVAATPTSLLISWSLDNQAKYYRITYGETGGNSLVQEFTMPGVTNATISGLKPGVDYTIT
VYAVTRVGRMLDTPGPISTNYRTEIDKPSQGS

L5.6.2 (16 ± 6 pM)
GCTAGCGCTCCTGATGTTCCGAGGGACCTGGAAGTTGTCGCTGCGGCCCCCACCAGCCTACTGATCAGCTG
GCGCGACTGCCCCTGGGCTATCTATTACGGGATCACTTACGGAGAAACAGGAGGAAATAGCCTTGTCCAGG
AGTTCACTATGCCCGGGGTTACCAATGCTACCATCAGCGGCCTTAAACCTGGAGTTGATTATACCATCACT
GTGTATGCTGTCACTCGCGTGGGGCGGATGTCTTGCACGTCGGGCCCAATTTCCATTAATTACCGAACAGA
AATTGACAAACCATCCCAGGGATCC

ASAPDVPRDLEVVAAAPTSLLISWRDCPWAIYYGITYGETGGNSLVQEFTMPGVTNATISGLKPGVDYTIT
VYAVTRVGRMSCTSGPISINYRTEIDKPSQGS

L6.5.1 (6.6 ± 1.3 pM)
GCTAGCGTTTCTGATGTTCCGAGGGACCTGGAAGTTGTTGCTGCGACCTCCACCAGCCTACTGATCAGCTG
GCGCGGCTGCCCCTGGGCTATCTATTACGGGATCACTTACGGAGAAACAGGAGGGAGTAGCCTTGTCCAGG
AGTTCACTATGCCTGGGGTTACCAATGCTACCATCAGCGGCCTTGAACCTGGAGTTGATTATACCATCACT

GTGTACGCTGTCACTCGCGTGGGGCGGATGCTTTGCACGCCGGGCCCAATTTCCATTAATTACCGAACAGA
AATTGACAGACCATCCCAGGGATCC

ASVSDVPRDLEVVAATSTSLLISWRGCPWAIYYGITYGETGGSSLVQEFTMPGVTNATISGLEPGVDYTIT
VYAVTRVGRMLCTPGPISINYRTEIDRPSQGS

L7.5.1 (2.6 ± 0.6 pM)
GCTAGCGTTTCTGATGTTCCGAGGGACCTGGAAGTTGTTGCTGCGACCTCCACCAGCCTACTGATCAGTTG
GCGCGGCTACCCCTGGGCTACCTATTATGGGATCATTTACGGAGAAACGGGAGGAAATAGCCTTGTCCAGG
AGTTCACTATGCCTGGGGTTACCAATGCTACCATCAGCGGCCTTAAACCTGGAGTTGATTATACCATCACT
GTGTATGCTGTCACTCGCGTGGGGCGGACGTTTGACACGCCGGGCCCAATCTCCATTAATTACCGAACAGA
AATTGACAAACCATCCCAGGGATCC

ASVSDVPRDLEVVAATSTSLLISWRGYPWATYYGIIYGETGGNSLVQEFTMPGVTNATISGLKPGVDYTIT
VYAVTRVGRTFDTPGPISINYRTEIDKPSQGS

L8.5.2 (2.9 ± 1.4 pM)
GCTAGCGTTTCTGATGTTCCGAGGGACCTGGAAGTTGTTGCTGCGACCTCCACCAGCCTACTGATCAGCTG
GCGCGGCTGCCCCTGGGCTATCTATTACGGGGTCACTTACGGAGAAACAGGAGGGAGTAGCCTTGCCCAGG
AGTTCACTATGCCTGGGGTTACCAATGCTACCATCAGCGGCCTTGAACCTGGAGTTGATTATACCATCACT
GTGTACGCTGTCACTCGCGTGGGGCGGATGCTTTGCGCGCCGGGCCCAATTTCCATTAATTACCGAACAGA
AATTGACAGACCATCCCAGGGATCC

ASVSDVPRDLEVVAATSTSLLISWRGCPWAIYYGVTYGETGGSSLAQEFTMPGVTNATISGLEPGVDYTIT
VYAVTRVGRMLCAPGPISINYRTEIDRPSQGS

L8.5.7 (2.8 ± 0.5 pM)
GCTAGCGTTTCTGGTGTTCCGAGGGACCTGGAAGTCGTTGCGGCGACCCCCACCAGCATACTGATCAGCTG
GCGCGACCGTCCCTGGGCTATCTATTACGGGATCACTTACGGAGAAACAGGAGGGAGTAGCCTTGTCCAGG
AGTTCACTATGCCTGGGGTTACCAATGCTACCATCAGCGGCCTTAAACCTGGAGTTGATTATACCATCACT
GTGCATGCTGTCACTAGGCTGTCCATTGTGCCATACGCCCCAATTTCCATTACTTACCGAACAGAAATTGA
CAAGCCACCCCAGGGATCC

ASVSGVPRDLEVVAATPTSILISWRDRPWAIYYGITYGETGGSSLVQEFTMPGVTNATISGLKPGVDYTIT
VHAVTRLSIVPYAPISITYRTEIDKPPQGS

Cons0.4.1 (1.1 ± 0.6 pM)
GCTAGCGTTTCTGATGTTCCGAGGGACCTGGAAGTTGTTGCTGCGACCCCCACCAGCCTACTGATCAGTTG
GCGCGAGGACCCCTGGGCTAAGTATTATGGGATCATTTACGGAGAAACGGGAGGAAATAGCCTTGTCCAGG
AGTTCACTATGCCTGGGGTTACCAATGCTACCATCAGCGGCCTTAAACCTGGAGTTGATTATACCATCACT
GTGTATGCTGTCACTAGAGTAGGTTGGGCGTCTTATACACTAGGCCCAATCTCCATTAATTACCGAACAGA
AATTGACAAACCATCCCAGGGATCC

ASVSDVPRDLEVVAATPTSLLISWREDPWAKYYGIIYGETGGNSLVQEFTMPGVTNATISGLKPGVDYTIT
VYAVTRVGWASYTLGPISINYRTEIDKPSQGS

Goat IgG
gI2.5.3T88I (1.2 ± 0.4 nM)
GCTAGCGTTTCTGATGTTCCGAGGGACCTGGAAGTTGTTGCTGCGACCCCCACCAGCCTACTGATCAGCTG
GACCCGTGCTTATTTTGCTCCGTATTACAGGATCACTTACGGAGAAACAGGAGGAAATAGCCCTGTCCAGG
AGTTCACTGTGTCTGGTTCGCTTTCCAGCGCTATCATCAGCGGCCTTAAACCTGGAGTTGATTATACCATC
ACTGTGTATGCTGTCACTTCCTACGGCCTCGTTATCACCGATCCAATTTCCATTAATTACCGAACAGAAAT
TGACAAACCATCCCAGGGATCC

ASVSDVPRDLEVVAATPTSLLISWTRAYFAPYYRITYGETGGNSPVQEFTVSGSLSSAIISGLKPGVDYTI
TVYAVTSYGLVITDPISINYRTEIDKPSQGS

gI2.5.2 (32 ± 21 nM)
GCTAGCGTTTCTGATGTTCCGAGGGACCTGGAAGTTGTTGCTGCGACCCCCACCAGCCTACTGATCAGCTG
GGCCTTGCCGCGGTCGGAGTATTACAGGATCACTTACGGAGAAACAGGGGGAAATAGCCCTGTCCAGGAGT
TCACTGTGCCTAATTGGACGTCTGCTATCATCAGCGGCCTTAAGCCTGGAGTTGATTATACCATCACTGTG

TATGCTGTCACTTCTCCTGGGTTGGTTCTGGGGGCGCCAATTTCCATTAATTACCGAACAGAAATTGACAA
ACCATCCCAGGGATCC

ASVSDVPRDLEVVAATPTSLLISWALPRSEYYRITYGETGGNSPVQEFTVPNWTSAIISGLKPGVDYTITV
YAVTSPGLVLGAPISINYRTEIDKPSQGS

gI2.5.4 (35 ± 16 nM)
GCTAGCGTTTCTGATGTTCCGAGGGACCTGGAAGTTGTTGCTGCGACCCCCACCAGCCTACTGATCAGCTG
GAAGATGCGTGCTGCTCGTTATTACAGGATCACTTACGGAGAAACAGGAGGAAATAGCCCTGTCCAGGAGT
TCACTGTGCNNAATTGGACGTCTGCTATCATCAGCGGCCTTAAACCTGGAGTTGATTATACCATCACTGTG
TATGCTGTCACTTCTCCTGGGTTGATTCTGGGGGCGCCAATTTCCATTAATTACCGAACAGGAATTGACAA
ACCATCCCAGGGATCC

ASVSDVPRDLEVVAATPTSLLISWKMRAARYYRITYGETGGNSPVQEFTV?NWTSAIISGLKPGVDYTITV
YAVTSPGLILGAPISINYRTGIDKPSQGS

Rabbit IgG

rI4.5.5 (51 ± 4 pM)
GCTAGCGTTTCTGATGTTCCGAGGGACCTGGAAGTTGTTGCTGCGACCCCCACCAGCCTACTGATCAGCTG
GGCGACTACCGGGAAGGCTCCCCTTTATTACAGGATCACTTACGGAGAAACAGGAGGAAATAGCCCTGTCC
AGGAGTTCACTGTGCCTGCTACGTGGGTCAAGGCTACCATCCGCGGCCTTAAACCTGGAGTTGATTATACC
ATCACTGTGTATGCTGTCACTCATTATGACGATACCCTGTCTCCAATTTCCATTAATTACCGAACAGAAAT
TGACAAACCATCCCAGGGATCC

ASVSDVPRDLEVVAATPTSLLISWATTGKAPLYYRITYGETGGNSPVQEFTVPATWVKATIRGLKPGVDYT
ITVYAVTHYDDTLSPISINYRTEIDKPSQGS

rI4.3.1 (117 ± 6 pM)
GCTAGCGTTTCTGATGTTCCGAGGGACCTGGAAGTTGTTGCTGCGACCCCCACCAGCCTACTGATCAGCTG
GGCGACTACCGGGAAGACTCCCCTTTATTACAGGATCACTTACGGAGAAACAGGAGGAAATAGCCCTGTCC
AGGAGTTCACTGTGCCTCGTTCTGCCGAGATGGCTACCATCAGCGGCCTTAAACCTGGAGTTGATTATACC
ATCACTGTGTATGCTGTCACTCACTATGACGATACCCTGTCTCCAATTTCCATTAATTACCGAACAGAAAT
TGACAAACCATCCCAGGGATCC

ASVSDVPRDLEVVAATPTSLLISWATTGKTPLYYRITYGETGGNSPVQEFTVPRSAEMATISGLKPGVDYT
ITVYAVTHYDDTLSPISINYRTEIDKPSQGS

rI4.3.4 (300 ± 120 pM)
GCTAGCGTTTCTGATGTTCCGAGGGACCTGGAAGTTGTTGCTGCGACCCCCACCAGCCTACTGATCAGCTG
GGTTAATGGGGACTCTTGTCTTTATTACAGGATCACTTACGGAGAAACAGGAGGAAATAGCCCTGTCCAGG
AGTTCACTGTGTCTGGTTCTGCTCATGTTGCTACCATCAGGCGCCTTAAACCTGGAGTTGATTATACCATC
ACTGTGTATACTGTCACTGGGTACGGTGGAAAGAGGGTGCAGCCAATTTCCATTAATTACCGAACAGAAAT
TGACAAACCATCCCAGGGATCC

ASVSDVPRDLEVVAATPTSLLISWVNGDSCLYYRITYGETGGNSPVQEFTVSGSAHVATIRRLKPGVDYTI
TVYTVTGYGGKRVQPISINYRTEIDKPSQGS

rI3.6.4 (0.63 ± 0.07 nM)
GCTAGCGTTTCTGATGTTCCGAGGGACCTGGAAGTTGTTGCTGCGACCCCCACCAGCCTACTGATCAGCTG
GGTTCGTCCCTCGTACAGTCGGTTGTATTACAGGATCACTTACGGAGAAACAGGAGGAAATAGCCCTGTCC
AGGAGTTCACTGTGTCTCGTTCGGCTCGCTCGGCTACCATCCGCGGCCTTAAACCTGGAGTTGATTATACC
ATCACTGTGTATGCTGTCACTGGGTACGGTGGGGAGAGGGTGCAGCCAATTTCCATTAATTACCGAACAGA
AATTGACAAACCATCCCAGGGATCC

ASVSDVPRDLEVVAATPTSLLISWVRPSYSRLYYRITYGETGGNSPVQEFTVSRSARSATIRGLKPGVDYT
ITVYAVTGYGGERVQPISINYRTEIDKPSQGS

rI4.3.3 (1.08 ± 0.38 nM)
GCTAGCGTTTCTGATGTTCCGAGGGACCTGGAAGTTGTTGCTGCGACCCCCACCAGCCTACTGATCAGCTG
GGCTCGTGCTTCGAATCCCTGTCTCTATTACAGGATCACTTACGGAGAAACAGGAGGAAATAGCCCTGTCC
AGGAGTTCACTGTGCCTGCTACGTGGGTCAAGGCTACCATCCGCGGCCTTAAACCTGGAGTTGATTATACC
ATCACTGTGTATGCTGTCACTGGGTACGGTGGGAAGAGGGTGCAGCCAATTTCCATTAATTACCGAACAGA
GATTGACAAACCATCCCAGGGATCC

ASVSDVPRDLEVVAATPTSLLISWARASNPCLYYRITYGETGGNSPVQEFTVPATWVKATIRGLKPGVDYT
ITVYAVTGYGGKRVQPISINYRTEIDKPSQGS

CEA

C7.4.3 (1.8 ± 0.4 nM)
GCTAGCGTTTCTGATGGTACTTTAAGCCGGGACCTGGGAGTTGTTGCTGCAACCCCCACCAGCCTACTGAT
CAGCTGGTATTACTCTTATTCTCATCACTACTCTTCTTACAGGATCACTTACGGAGAAACAGGAGGAAATA
GCCCTGTCCAGGAGTTCACTGTGCCTAGGTATCGGGCCTTTGCTACCATCAGCGGCCTTAAACCTGGAGTT
GATTATACCATCACTGTGTATGCTGTCACTTCTTCTTCCTCTTACTCCTATCCAATTTCCATTAATTACCG
AACAGAAATTGACAAACCATCCCAGGGATCC

ASVSDGTLSRDLGVVAATPTSLLISWYYSYSHHYSSYRITYGETGGNSPVQEFTVPRYRAFATISGLKPGV
DYTITVYAVTSSSSYSYPISINYRTEIDKPSQGS

NNB C3.2.1 (~39 nM)
GCTAGCGTTTCTGATGTTCCGAGGGACCTGGAAGTTGTTGCTGCGACCCCCACCAGCCTACTGATCAGCTG
GAGGCATGTGCGGGAGCATTATTACAGGGTCACTTACGGAGAAACAGGAGGAAATAGCCCTGTCCAGGGGT
TCACTGTGCCTCCGCGCCTTGGTCGTGCTACCATCGGCGGCCTTAAACCTGGAGTTGATTATACCATCACT
GTGTATGCTGTCACTTTGGGTCCCCATGTGCCAATTTCCATTAATTACCGAACAGAAATTGACAAACCATC
CCAGGGATCC

ASVSDVPRDLEVVAATPTSLLISWRHVREHYYRVTYGETGGNSPVQGFTVPPRLGRATIGGLKPGVDYTIT
VYAVTLGPHVPISINYRTEIDKPSQGS

Mouse IgG

mI3.2.1 (4.1 ± 0.7 nM)
GCTAGCGTTTCTGATGTTCCGAGGGACCTGGAAGTTGTTGCTGCGACCCCCACCAGCCTACTGATCAGCTG
GTGTTGCTCCGATAACTGTTCAAATTCTTACAGGATCACTTACGGAGAAACAGGAGGAAATAGCCCTGTCC
AGGAGTTCACTGTGCCTCGCTCGTGCTTCATGGCTACCATCAGCGGCCTTAAACCTGGAGTTGATTATACC
ATCACTGCGTATGCTGTCACTGATAGTAACGGTCCTCATCCAATTTCCATTAATTACCGAACAGAAATTGA
CAAACCATCCCAGGGATCC

ASVSDVPRDLEVVAATPTSLLISWCCSDNCSNSYRITYGETGGNSPVQEFTVPRSCFMATISGLKPGVDYT
ITAYAVTDSNGPHPISINYRTEIDKPSQGS

Albumin

Alb3.2.1
GCTAGCGTTTCTGATGTTCCGAGGGACCTGGAAGTTGTTGCTGCGACCCCCACCAGCCTACTGATCAGTTG
GCGCGGCTACCCCTGGGCTACCTATTATGGGATCACTTACGGAGAAACAGGAGGAAATAGCCTTGTCCAGG
AGTTCACTATGCCTGGGGTTACCAATGCTACCATCAGCGGCCTTAAACCTGGAGTTGATTATACCATCACT
GTGTATGCTGTCACTCGCGTGGGGCGGACGTTTGACACGCCGGGCCCAATCTCCATTAATTACCGAACAGA
AATTGACAAACCATCCCAGGGATCC

ASVSDVPRDLEVVAATPTSLLISWRGYPWATYYGITYGETGGNSLVQEFTMPGVTNATISGLKPGVDYTIT
VYAVTRVGRTFDTPGPISINYRTEIDKPSQGS

EGFR

E6.2.6 (260 ± 130 pM) [6.2.6, Clone A]
GCTAGCGTTTCCGATGTTCCGAGGGACCTGGAGGTTGTTGCTGCGACCCCCACCAGCCTACTGATCAGCTG
GTTCGACTACGCTGTGACTTATTACAGGATCACTTACGGAGAAACAGGAGGAAATAGCCCTGTCCAGGAGT
TCACTGTGCCTGGTTGGATCTCCACTGCTACCATCAGCGGCCTTAAACCTGGAGTTGATTATACCATCACT
GTGTATGCTGTCACTGACAACTCTCGTTGGCCTTTTCGCTCTACTCCAATTTCCACTAATTACCGAACAGA
AATTGACAAACCACCCCAGGGATCC

ASVSDVPRDLEVVAATPTSLLISWFDYAVTYYRITYGETGGNSPVQEFTVPGWISTATISGLKPGVDYTIT
VYAVTDNSRWPFRSTPISTNYRTEIDKPPQGS

E4.2.1 (250 ± 70 pM) [B2, Clone B]

GCTAGCGTTTCTGATGTTCCGAGGGACCTGGAAGTTGTTGCTGCGACCCCCACCAGCCTACTGATCAGCTG
GTACGGTTTTTCGCTTGCGAGCTCTTACAGGATCACTTACGGAGAAACAGGAGGAAATAGCCCTGTCCAGG
AGTTCACTGTGCCTCGTTCGCCCTGGTTTGCTACCATCAGCGGCCTTAAACCTGGAGTTGATTATACCATC
ACTGTGTATGCTGTCACTTCTAACGACTTTTCTAATCGTTACTCTGGTCCAATTTCCATTAATTACCGAAC
AGAAATTGACAAACCATCCCAGGGATCC

ASVSDVPRDLEVVAATPTSLLISWYGFSLASSYRITYGETGGNSPVQEFTVPRSPWFATISGLKPGVDYTI
TVYAVTSNDFSNRYSGPISINYRTEIDKPSQGS

EI4.4.2 (1.4 ± 0.2 nM) [U2, Clone C]
GCTAGCGTTTCTGATGTTCCGAGGGACCTGGAAGTTGTTGCTGCGACCCCCACCAGCCTACTGATCAGCTG
GTATTTTCGCGACCCCCGGTACGTGGACTATTACAGGATCACTTACGGAGAAACAGGAGGAAATAGCCCTG
CCCAGGAGTTCACTGTGCCTTGGTACCTTCCTGAGGCTACCATCAGCGGCCTTAAACCCGGAGTTGATTAT
ACCATCACTGTGTATGCTGTCACTGGGGACGATCAGAATGCTGGGCTTCCAATTTCCATTAATTACCGAAC
AGAAATTGACAAACCATCCCAGGGATCC

ASVSDVPRDLEVVAATPTSLLISWYFRDPRYVDYYRITYGETGGNSPAQEFTVPWYLPEATISGLKPGVDY
TITVYAVTGDDQNAGLPISINYRTEIDKPSQGS

EI3.4.3 (250 ± 50 pM) [U3, Clone D]
GCTAGCGTTTCTGATGTTCCGAGGGACCTGGAAGTTGTTGCTGCGACCCCCACCAGCCTACTGATCAGCTG
GCTTCACCATCGCTCTGACGTGCGCTCTTACAGGATCACTTACGGAGAAACAGGAGGAAATAGCCCTGTCC
AGAAGTTCACTGTGCCTGGGTCGCGCTCCCTGGCTACCATCAGCGGCCTTAAACCTGGAGTTGATTATACC
ATCACTGTGTATGCTGTCACTTGGGGGTCTTACTGTTGCTCTAATCCAATTTCCATTAATTACCGAACAGA
AATTGACAAACCATCCCAGGGATCC

ASVSDVPRDLEVVAATPTSLLISWLHHRSDVRSYRITYGETGGNSPVQKFTVPGSRSLATISGLKPGVDYT
ITVYAVTWGSYCCSNPISINYRTEIDKPSQGS

EI2.4.6 (2.9 ± 0.3 nM) [U4, Clone E]

GCTAGCGTTTCTGATGTTCCGAGGGACCTGGAAGTTGTTGCTGCGACCCCCACCAGCCTACTGATCAGCTG
GTACCTTCGTGACCCCCGGTACGTGGACTATTACAGGATCACTTACGGAGAAACAGGAGGAAATAGCCCTG
TCCAGGAGTTCACTGTGCCTTGGTACCTTCCTGAGGCTACCATCAGCGGCCTTAAACCTGGAGTTGATTAT
ACCATCACTGTGTATGCTGTCACTTACGATGGCTACCGCGAGAGTACCCCTCTCCCAATTTCCATTAATTA
CCGAACAGAAATTGACAAACCATCCCAGGGATCC

ASVSDVPRDLEVVAATPTSLLISWYLRDPRYVDYYRITYGETGGNSPVQEFTVPWYLPEATISGLKPGVDY
TITVYAVTYDGYRESTPLPISINYRTEIDKPSQGS

EI3.4.2 (9.5 ± 3.5 nM) [FG5]
GCTAGCGTTTCTGATGTTCCGAGGGACCTGGAAGTTGTTGCTGCGACCCCCACCAGCCTACTGATCAGCTG
GTATGGTTCCAGTTACGCGTCCTATTACAGGATCACTTACGGAGAAACAGGAGGAAATAGCCCTGTCCAGG
AGTTCACTGTGCCTCGTTCGCCCTGGTTTGCTATCATCAGCGGCCTGAAACCTGGAGTTGATTATACCATC
ACTGTGTATGCTGTCACTCCTAGTGGGATCTCTGCTCCAATTTCCATTAATTACCGAACAGAAATTGACAA
ACCATCCCAGGGATCC

ASVSDVPRDLEVVAATPTSLLISWYGSSYASYYRITYGETGGNSPVQEFTVPRSPWFAIISGLKPGVDYTI
TVYAVTPSGISAPISINYRTEIDKPSQGS

EI1.4.1 (0.85 ± 0.50 nM) [U5]
GCTAGCGTTTCTGATGTTCCGAGGGACCTGGAAGTTGTTGCTGCGACCCCCACCAGCCTACTGATCAGCTG
GTATCATCCTTTCTATTATGTCGCGCATTCTTACAGGATCACTTACGGAGAAACAGGAGGAAATAGCCCTG
TCCAGGAGTTCACTGTGCCTCGTTCGCCCTGGTTTGCTACCATCAGCGGCCTTAAACCTGGAGTTGATTAT
ACCATCACTGTGTATGCTGTCACTAGTAAGTGCTATGATGGTTCTGTCCCAATTTCCATTAATTACCGAAC
AGAAATTGACAAACCATCCCAGGGATCC

ASVSDVPRDLEVVAATPTSLLISWYHPFYYVAHSYRITYGETGGNSPVQEFTVPRSPWFATISGLKPGVDY
TITVYAVTSKCYDGSVPISINYRTEIDKPSQGS

CD276
CD6.3.1 (54 ± 38 pM)

GCTAGCGTTTCTGATGTCCCGAGGGACCTGGAAGTTGTTGCTGCGACCCCCACCAGCCTACTGATCAGCTG
GTTCGGCTATTACGGCGCGCTTTTACAGGATCACTTACGGGGAAACAGGAGGAAATAGCCCTGTCCAGG
AGTTCACTGTGCCTGGGCGCTTTTCCAGCTACACCGCCACCATCAGCGGCCTTAAACCTGGAGTTGATTAT
ACCATCACTGTGTATGCTGTCACTGATAATGTTGGGTCTTATCCAATTCCCATTAATTACCGAACAGAAAT
TGACAAACCATCCCAGGGATCC

ASVSDVPRDLEVVAATPTSLLISWFGYYGARFYRITYGETGGNSPVQEFTVPGRFSSYTATISGLKPGVDY
TITVYAVTDNVGSYPIPINYRTEIDKPSQGS

CD6.3.2 (17 ± 14 pM)
GCTAGCGTTTCTGATGTTCCGAGGGACCTGGAAGTTGTTGCTGCGACCCCCACCAGCCTACTGATCAGCTG
GTCTTATCCCTGTCTTTTTCAGGTGCACTATTACAGGATCACTTACGGAGAAACAGGAGGAAATAGCCCTG
TCCAGGAGTTCACTGTGCCTGGGGGCTTTTCCGGCTACACCGCTACCATCAGCGGCCTTAAACCTGGAGTT
GATTATACCATCACTGTGTATGCTGTCACTGATTACTCTTTCCATCACGATTGCTCTTCTCCAATTTCCAT
TAATTACCGAACAGAAATTGACAAACCATCCCAGGGATCC

ASVSDVPRDLEVVAATPTSLLISWSYPCLFQVHYYRITYGETGGNSPVQEFTVPGGFSGYTATISGLKPGV
DYTITVYAVTDYSFHHDCSSPISINYRTEIDKPSQGS

CD6.3.6
GCTAGCGTTTCTGATGTTCCGAGGGACCTGGAAGTTGTTGCTGCGACCCCCACCAGCCTACTGATCAGCTG
GTACTATCCTGACTACACGTTGTATTACAGGATCACTTACGGAGAAACAGGAGGAAATAGCCCTGTCCAGG
AGTTCACTGTGCCTGGGCGCTTTCCCGGCTACACCGCTACCATCAGCGGCCTTAAACCTGGAGTTGATTAT
ACCATCACTGTGTATGCTGTCACTGGTTACCGTGTCTACGACCGCTACTCTCATCCAATTTCCATTAATTA
CCGAACAGAAATTGACAAACCATCCCAGGGATCC

ASVSDVPRDLEVVAATPTSLLISWYYPDYTLYYRITYGETGGNSPVQEFTVPGRFPGYTATISGLKPGVDY
TITVYAVTGYRVYDRYSHPISINYRTEIDKPSQGS

CD6.3.8 (1.6 ± 0.9 pM)
GCTAGCGTTTCTGATGTTCCGAGGGACCTGGAAGTTGTTGCTGCGACCCCCACCGGCCTACTGATCAGCTG
GTCTTATCCCTGTCTTTTTCGGGTGCACTATTACAGGATCACTTACGGAGAAACAGGAGGAAATAGCCCTG
TCCAGGAGTTCACTGTGCCTGGGGGCTTTTCCGGCTACACCGCTACCATCAGCGGCCTTAAACCTGGAGTT
GATTATACCATCACTGTGTATGCTGTCACTGGTTACTATTTCCGTCACGATTGCTCTTCGCCAATTTCCAT
TAAGTACCGAACAGAAATTGACAAACCATCCCAGGGATCC

ASVSDVPRDLEVVAATPTGLLISWSYPCLFRVHYYRITYGETGGNSPVQEFTVPGGFSGYTATISGLKPGV
DYTITVYAVTGYYFRHDCSSPISIKYRTEIDKPSQGS

Oligonucleotides

Name	Sequence	Use	Sites
a1	GACTGAGCTAGCGTTTCTGATGTTCCGAGGGACCTGGAAGTTGTTGCTGC	NNB library - framework	-5 to 38
a2	CCAGCTGATCAGTAGGCTGGTGGGGGTCGCAGCAACAACTTCCAGGTC	NNB library - framework	19 to 66
b3nm2	CACCAGCCTACTGATCAGCTGGNNBNNBNNBNNBNNBNNBTATTACAGGATCACTTACGGAG	NNB library - BC loop	45 to 112
b3nm1	CACCAGCCTACTGATCAGCTGGNNBNNBNNBNNBNNBNNBNNBTATTACAGGATCACTTACGGAG	NNB library - BC loop	45 to 112
b3n0	CACCAGCCTACTGATCAGCTGGNNBNNBNNBNNBNNBNNBNNBNNBTATTACAGGATCACTTACGGAG	NNB library - BC loop	45 to 112
b3np1	CACCAGCCTACTGATCAGCTGGNNBNNBNNBNNBNNBNNBNNBNNBNNBTATTACAGGATCACTTACGGAG	NNB library - BC loop	45 to 112
b4n	CACAGTGAACTCCTGGACAGGGCTATTTCCTCCTGTTTCTCCGTAAGTGATCCTGTAATA	NNB library - framework	91 to 150
c5nm1	CCCTGTCCAGGAGTTCACTGTGCCTNNBNNBNNBNNBGCTACCATCAGCGGCCTTAAACCTG	NNB library - DE loop	129 to 193
c5n0	CCCTGTCCAGGAGTTCACTGTGCCTNNBNNBNNBNNBNNBGCTACCATCAGCGGCCTTAAACCTG	NNB library - DE loop	129 to 193
c5np1	CCCTGTCCAGGAGTTCACTGTGCCTNNBNNBNNBNNBNNBNNBGCTACCATCAGCGGCCTTAAACCTG	NNB library - DE loop	129 to 193
c5np2	CCCTGTCCAGGAGTTCACTGTGCCTNNBNNBNNBNNBNNBNNBNNBGCTACCATCAGCGGCCTTAAACCTG	NNB library - DE loop	129 to 193
c6t	GTGACAGCATACACAGTGATGGTATAATCAACTCCA	NNB library -framework	174 to 227
d7nm5	CCATCACTGTGTATGCTGTCACTNNBNNBNNBNNBNNBCCAATTTCCATTAATTACCGAA	NNB library - FG loop	206 to 280
d7nm4	CCATCACTGTGTATGCTGTCACTNNBNNBNNBNNBNNBNNBCCAATTTCCATTAATTACCGAA	NNB library - FG loop	206 to 280
d7nm2	CCATCACTGTGTATGCTGTCACTNNBNNBNNBNNBNNBNNBNNBNNBCCAATTTCCATTAATTACCGAA	NNB library - FG loop	206 to 280
d7n0	CCATCACTGTGTATGCTGTCACTNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBCCAATTTCCATTAATTACCGAA	NNB library - FG loop	206 to 280
d8n	GGATCCCTGGGATGGTTTGTCAATTTCTGTTCGGTAATTAATGGAAATTGG	NNB library -framework	259 to 309
p1	GACTGAGCTAGCGTTTCTGATG	NNB library - amplify	-
p8	CGTCATGGATCCCTGGGATG	NNB library - amplify	-
a1YS	GGTTCTGCTAGCGTTTCTGATGTTCCGAGGGACCTGGAAGTTGTTGCTGC	YS library - framework	-5 to 38
b3YSa	CCAGCCTACTGATCAGCTGGTMYTMYTMYTMYTMYTMYTATTACAGGATCACTTACGGAG	YS library - BC loop	47 to 112
b3YSb	CCAGCCTACTGATCAGCTGGTMYTMYTMYTMYTMYTMYTMYTATTACAGGATCACTTACGGAG	YS library - BC loop	47 to 112
b3YSc	CCAGCCTACTGATCAGCTGGTMYTMYTMYTMYTMYTMYTMYTMYTATTACAGGATCACTTACGGAG	YS library - BC loop	47 to 112
b3YSd	CCAGCCTACTGATCAGCTGGTMYTMYTMYTMYTMYTMYTMYTMYTMYTATTACAGGATCACTTACGGAG	YS library - BC loop	47 to 112
c5YSa	CTGTCCAGGAGTTCACTGTGCCTggBtcBtcBacBGCTACCATCAGCGGCCTTAAAC	YS library - DE loop	131 to 190
c5YSb	CTGTCCAGGAGTTCACTGTGCCTggBtcBNNBtcBacBGCTACCATCAGCGGCCTTAAAC	YS library - DE loop	131 to 190
c5YSc	CTGTCCAGGAGTTCACTGTGCCTggBtcBNNBNNBNNBtcBacBGCTACCATCAGCGGCCTTAAAC	YS library - DE loop	131 to 190
c6YS	GTGACAGCATACACAGTGATGGTATAATCAACTCCAGGTTTAAGGCCGCTGATGGTAGC	YS library - framework	169 to 227
d7YSa	CATCACTGTGTATGCTGTCACTTMYTMYTMYTMYTMYTMYCCAATTTCCAGAAATTGACAAAC	YS library - FG loop	207 to 281
d7YSb	CATCACTGTGTATGCTGTCACTTMYTMYTMYTMYTMYTMYTMYCCAATTTCCAGAAATTGACAAAC	YS library - FG loop	207 to 281
d7YSc	CATCACTGTGTATGCTGTCACTTMYTMYTMYTMYTMYTMYTMYTMYCCAATTTCCAGAAATTGACAAAC	YS library - FG loop	207 to 281
d7YSd	CATCACTGTGTATGCTGTCACTTMYTMYTMYTMYTMYTMYTMYTMYTMYTMYCCAATTTCCAGAAATTGACAAAC	YS library - FG loop	207 to 281
p1YS	GGTTCTGCTAGCGTTTCTGATGTTCCGAGG	YS library - amplify	-
p8YS	TTGTTCGGATCCCTGGGATGGTTTGTCAATTTC	YS library - amplify	-
b3G4m2	CCAGCCTACTGATCAGCTGG(15152545)(45152515)(00451045)(15152545)(45301510)(00451045)(05650525)(20650510)(00451045)(15152545)(45152515)(00451045)KYG(15152545)(45152515)(00451045)TMTTACAGGATCACTTACGGAGAAAC	G4 library - BC loop	47 to 116
b3G4m1	CCAGCCTACTGATCAGCTGG(15152545)(45152515)(00451045)(15152545)(45301510)(00451045)(05650525)(20650510)(00451045)(15152545)(45152515)(00451045)(15152545)(45152515)(00451045)KYG(15152545)(45152515)(00451045)TMTTACAGGATCACTTACGGAGAAAC	G4 library - BC loop	47 to 116
b3G40	CCAGCCTACTGATCAGCTGG(15152545)(45152515)(00451045)(15152545)(45301510)(00451045)(05650525)(20650510)(00451045)(15152545)(45152515)(00451045)(15152545)(45152515)(00451045)(15152545)(45152515)(00451045)KYG(15152545)(45152515)(00451045)TMTTACAGGATCACTTACGGAGAAAC	G4 library - BC loop	47 to 116
b3G4p1	CCAGCCTACTGATCAGCTGG(15152545)(45152515)(00451045)(15152545)(45301510)(00451045)(05650525)(20650510)(00451045)(15152545)(45152515)(00451045)(15152545)(45152515)(00451045)(15152545)(45152515)(00451045)(15152545)(45152515)(00451045)KYG(15152545)(45152515)(00451045)TMTTACAGGATCACTTACGGAGAAAC	G4 library - BC loop	47 to 116
b4G4	GCACAGTGAACTCCTGGACAGGGCTATTTCCTCCTGTTTCTCCGTAAGTGATCCTG	G4 library - framework	96 to 151
d7G4m4	CATCACTGTGTATGCTGTCACT(15103540)(40103515)(00451045)(15152545)(45152515)(00451045)(15103540)(40103515)(00451045)(15152545)(45152515)(00451045)TCT(15152545)(45152515)(00451045)CCAATTTCCATTAATTACCGAAC	G4 library - FG loop	207 to 281
d7G4m3	CATCACTGTGTATGCTGTCACT(15103540)(40103515)(00451045)(15152545)(45152515)(00451045)(15103540)(40103515)(00451045)(15152545)(45152515)(00451045)(15152545)(45152515)(00451045)TCT(15152545)(45152515)(00451045)CCAATTTCCATTAATTACCGAAC	G4 library - FG loop	207 to 281
d7G4m2	CATCACTGTGTATGCTGTCACT(15103540)(40103515)(00451045)(15152545)(45152515)(00451045)(15103540)(40103515)(00451045)(15152545)(45152515)(00451045)(15152545)(45152515)(00451045)(15152545)(45152515)(00451045)TCT(15152545)(45152515)(00451045)CCAATTTCCATTAATTACCGAAC	G4 library - FG loop	207 to 281
g7G40	CATCACTGTGTATGCTGTCACT(15103540)(40103515)(00451045)(15152545)(45152515)(00451045)(15103540)(40103515)(00451045)(15152545)(45152515)(00451045)(15152545)(45152515)(00451045)(15152545)(45152515)(00451045)(15152545)(45152515)(00451045)TCT(15152545)(45152515)(00451045)CCAATTTCCATTAATTACCGAAC	G4 library - FG loop	207 to 281
W5	CGACGATTGAAGGTAGATACCCATACGACGTTCCAGACTACGCTCTGCAG	gene mutagenesis	-
W3	ATCTCGAGCTATTACAAGTCCTCTTCAGAAATAAGCTTTTGTTCGGATCC	gene mutagenesis	-
BC5	GATGTTCCGAGGGACCTGGAAGTTGTTGCTGCGACCCCCACCAGCCTACTGATCAGCTGG	BC loop mutagenesis	-
BC3	TGAACTCCTGGACAGGGCTATTTCCTCCTGTTTCTCCGTAAGTGATCCTGTAATA	BC loop mutagenesis	-
DE5	CAGGATCACTTACGGAGAAACAGGAGGAAATAGCCCTGTCCAGGAGTTCACTGTG	DE loop mutagenesis	-
DE3	GCATACACAGTGATGGTATAATCAACTCCAGGTTTAAGGCCGCTGATGGTAGC	DE loop mutagenesis	-
FG5	ACCATCAGCGGCCTTAAACCTGGAGTTGATTATACCATCACTGTGTATGCTGTC	FG loop mutagenesis	-
FG3	TCGAGCTATTACAAGTCCTCTTCAGAAATAAGCTTTTGTTCGGATCCCTGGGATGGTTTGTCAATTTCTGTTCGGTAATTAATGGAAATTGG	FG loop mutagenesis	-
GeneAmp5	CGACGATTGAAGGTAGATACCCATACG	gene amplification	-
GeneAmp3	ATCTCGAGCTATTACAAGTCCTCTTC	gene amplification	-
BCamp5	GATGTTCCGAGGGACCTGGAAGTTG	BC loop amplification	-
BCamp3	TGAACTCCTGGACAGGGCTATTTCC	BC loop amplification	-
DEamp5	CAGGATCACTTACGGAGAAACAGGAGG	DE loop amplification	-
DEamp3	GCATACACAGTGATGGTATAATCAAC	DE loop amplification	-
FGamp5	ACCATCAGCGGCCTTAAACCTGGAG	FG loop amplification	-
FGamp3	TCGAGCTATTACAAGTCCTCTTCAG	FG loop amplification	-
G4bc3	tgaactcctggacagggctatttcctcctgtttctccgtaagtgatcctgta	BC loop amplification	avoid Y31
G4de5	caggatcacttacggagaaacaggaggaaatagccctgtccaggagttcactgtgcct	DE loop amplification	include P51
G4fg5	accatcagcggccttaaacctggagttgattataccatcactgtgtatgctgtcact	FG loop amplifcation	include T76
b4DE	AGGCATAGTGAACTCCTGGAC	L7.5.1 S15P DE library	-
c5DEa	GTCCAGGAGTTCACTATGCCTNNBNNBNNBNNBGCTACCATCAGCGGCCTTAAA	L7.5.1 S15P DE library	-
c5DEb	GTCCAGGAGTTCACTATGCCTNNBNNBNNBNNBNNBGCTACCATCAGCGGCCTTAAA	L7.5.1 S15P DE library	-
c5DEc	GTCCAGGAGTTCACTATGCCTNNBNNBNNBNNBNNBNNBGCTACCATCAGCGGCCTTAAA	L7.5.1 S15P DE library	-
c5DEd	GTCCAGGAGTTCACTATGCCTNNBNNBNNBNNBNNBNNBNNBGCTACCATCAGCGGCCTTAAA	L7.5.1 S15P DE library	-
ConsBC	GCCTACTGATCAGTTGGCGCNNBNNBCCCTGGGCTNNBTATTATGGGATCATTTACGGAG	L7.5.1 S15P Cons library	-
ConsFG	GGTAATTAATGGAGATTGGGCCVNNVNNVNNVNNVNNCCGCCCCACGCGAGTGACAGCA	L7.5.1 S15P Cons library	-
ConsensusBC	GCCTACTGATCAGTTGGAGAGRTYRTCCTTGGGCTAYCTATTATGGGATCATTTACGGAG	L7.5.1 S15P Cons library	-
ConsensusFG	GGTAATTAATGGAGATTGGACYTRRTGYAYMARRCRYCCRACCTACTCTAGTGACAGCA	L7.5.1 S15P Cons library	-
Cons5	ACCTGGAAGTTGTTGCTGCGACCCCCACCAGCCTACTGATCAGTTGGCGC	L7.5.1 S15P Cons library	-
Cons3	CCCTGGGATGGTTTGTCAATTTCTGTTCGGTAATTAATGGAGATTGGGCC	L7.5.1 S15P Cons library	-
bcCEA	CACCAGCCTACTGATCAGCTGGBRTBRCYCTBRCVRTBRTBRCKDTVRCTCTTACAGGATCACTTACGGAG	CEA affinity maturation	45 to 112
fgCEA	TTCGGTAATTAATGGAAATTGGAHVGBBAHVABBGBBGBBABBAGTGACAGCATACACAGTGATG	CEA affinity maturation	207 to 280
atermmid	ATGTGTAAAGTTGGTAACGGAACG	pCT sequencing primer	-
pETfor	AAATTAATACGACTCACTATAGGG	pET sequencing primer	-

APPENDIXE. PROTOCOLS

Library Construction (demonstrated using G4 oligos)

Prepare Oligonucleotides

1. Add appropriate amount of 0.1x elution buffer to oligo to yield 100 uM solution.
2. Add 180 uL of 0.1x elution buffer to empty vials.
3. Add 20 uL of 100 uM oligo to yield 10 uM oligo.

Create Fn3 Library Fragments 1-4 and 5-8

1. Create a PCR mix:

61.2 uL	10x KOD buffer
61.2 uL	2 mM dNTPs
24.5 uL	25 mM $MgSO_4$
12.24 uL	KOD polymerase
153 uL	ddH$_2$O
244.8 uL	betaine (2.5M)
18.36 uL	DMSO

2. Make twelve 47 uL aliquots.
3. Add 1.0 uL of oligo 1 and 2.0 uL of oligo 2.

(1 uL)(10 pmol/uL)(6x10^{11} molec. / pmol) = 6x10^{12} molecules

Tube	Oligo 1	Oligo 2
1-2	a2	a1
3*b*-4 (a-d)	b3G4*b*	b4G4
5*d*-6 (a-c)	c6YS	c5YS*d*
7*f*-8 (a-d)	d7G4*f*	d8n

4. Thermocycle

95°	2 min.	
94°	30 s	10 cycles
58°	30 s	
68°	1 min.	
68°	10 min.	
4°	hold	

5. Combine 12 uL of products in vials.

6x10^{12} molecules x 12/50 = 1.44x10^{12} molecules

Tube	Tube 1	Tube 2	
1-4*b* (a-d)	1-2	3*b*-4	4 combinations
5-8*df* (a-c, a-d)	5*d*-6	7*f*-8	12 combinations

6. Run the same thermocycle.

Create Fn3 G4 Library Genes

At this point, small differences in DE loop length should not present significant differences in amplification efficiency or secondary structure of the 5-8 fragment.

1. Save 10 uL of PCR products for gel analysis or later library creation.
2. Add 135 uL of ddH$_2$O to remaining 15 uL

1.44x10^{12} x (15/25) / 150 uL = ~5x10^{9} molecules / uL

3. Create a PCR mix.

40.4 uL	10x KOD buffer
40.8 uL	2 mM dNTPs
16.2 uL	25 mM $MgSO_4$
8.1 uL	KOD polymerase
36.4 uL	ddH$_2$O
162 uL	betaine (2.5M)
12.1 uL	DMSO

4. Make eight 39 uL aliquots.
5. Add 1 uL of primer and 10 uL of template.

(1 uL)(10 pmol/uL)(6x10^{11} molec. / pmol) = 6x10^{12} molecules (from ~5x10^{10} molecules)

Tube	Primer	Template	
amp1-4*b*	p1YS	1-4*b*	4 reactions (*b* = a-d)
amp5-8*f*	p8YS	5-8*df*	4 reactions (*f* = a-d)
			DE loop: a 1.2 uL, b 6.9 uL, c 1.9 uL

6. Run the same thermocycle but with 15 cycles.
7. Combine 12 uL of product.

   (6x10^{12} molecules)(12 / 50) = 1.44x10^{12} molecules

Tube	Tube 1	Tube 2
YSgene*bf*	amp1-4*b*	amp5-8*f*

8. Run the same thermocycle (10 cycles).

   (1.5x10^{12} molecules)(run 12.5 uL / 25 uL total)(326 bp)(618 g/mol / bp)(mol / 6x10^{23} molec.) = 0.25 ug

9. Add 12.5 uL of product to 1.4 uL of 10x stop buffer.
10. Run DNA on 1.5% agarose gel.
11. Extract ~330 bp fragment (16 samples).

    (1.5x10^{12} molecules)(run 12.5 uL / 25 uL)(50% yield) / 40 uL = 9.4x10^{9} molecules/uL

12. Create a PCR mix.

81.6 uL	10x KOD buffer
81.6 uL	2 mM dNTPs
32.6 uL	25 mM $MgSO_4$
16.32 uL	KOD polymerase
40.8 uL	p1YS
40.8 uL	p8YS
326 uL	betaine (2.5M)
24.48 uL	DMSO

    (2.75 uL)(10 pmol / uL)(330 bp)(618 pg/pmol / bp)(1 ug / 10^{6} pg) = 5.6 ug

13. Make 16x40 uL aliquots.
14. Add 10 uL of gel extracted gene (9.4x10^{10} molecules).
15. Run the same thermocycle but with 35 cycles.
16. Concentrate with PelletPaint.

Transformation

1. Transform into EBY100 using electroporation and homologous recombination with pCT-Fn3-Loop.

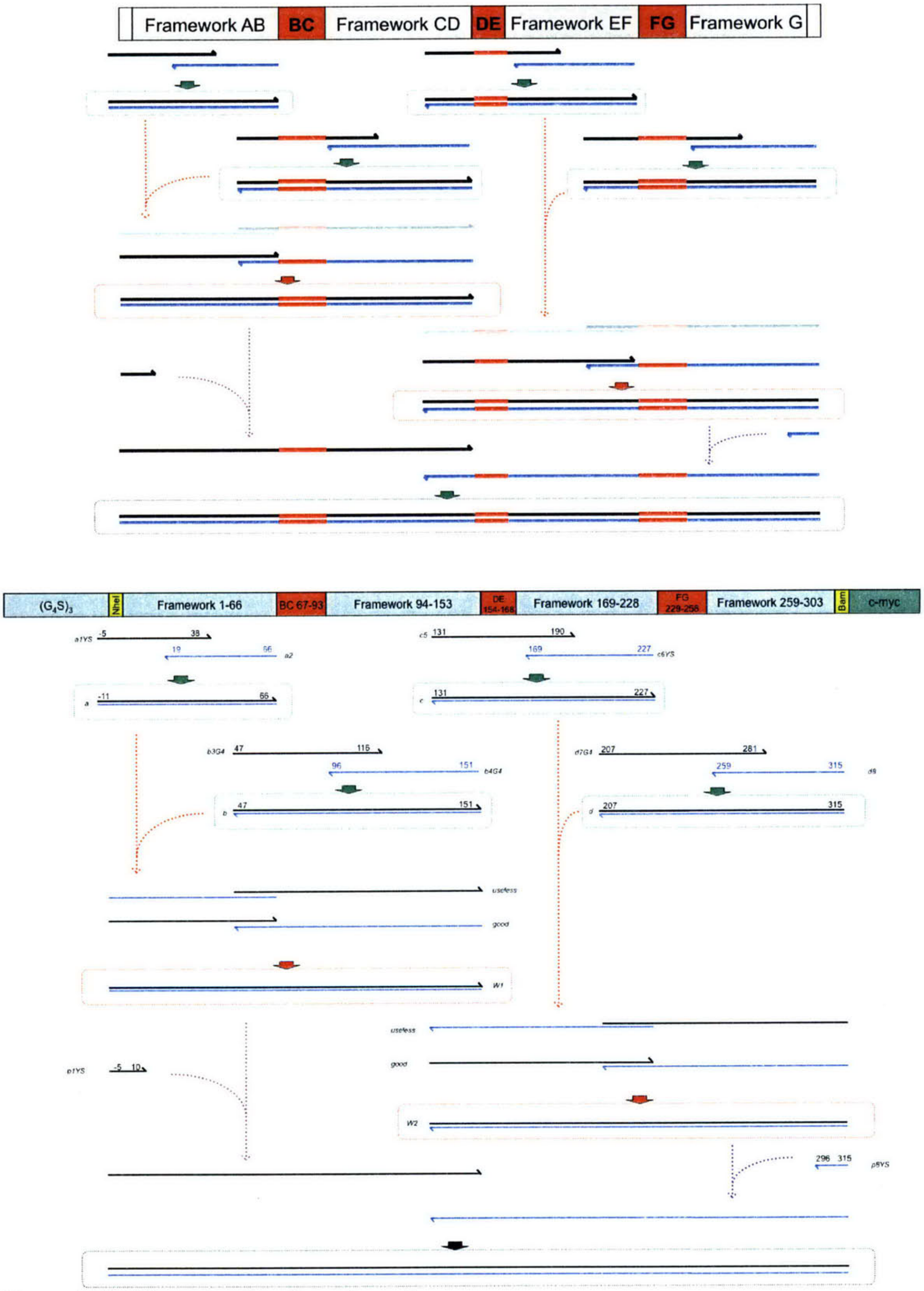

Framework AB
BC
Framework CD
DE
Framework EF
FG
Framework G
$(G_4S)_3$
NheI
Framework 1-66
BC 67-93
Framework 94-153
DE 154-168
Framework 169-228
FG 229-258
Framework 259-303
Bam
c-myc
a1YS
-5
38
19
66
a2
a
-11
66
c5
131
190
169
227
c6YS
c
131
227
b3G4
47
116
96
151
b4G4
b
47
151
d7G4
207
281
259
315
d8
d
207
315
useless
good
W1
useless
good
W2
p1YS
-5
10
296
315
p8YS

Figure E.1 *Library construction schematic.*

Magnetic Bead Sort - Initial

The naïve library will be depleted of streptavidin:bead binders and enriched for antigen binders. The diversity of the G4 library is $25x10^7$. Thus, $375x10^7$ cells (15x) will be sorted.

Protein Display

1. Grow and induce G4 library to display Fn3.
2. Use immediately or store yeast at 4°.
   Yeast displaying Fn3 can be used for several weeks if stored at 4°.

Bead Preparation

1. Combine 100 uL of PBSA + 6.7-33 pmoles of biotin-antigen + 10 uL of beads.
   Make two batches for initial sort.
   (10 uL beads)($4x10^5$ beads/uL)($1x10^6$ Ag/bead)(pmol / $6x10^{11}$ Ag) = 6.7 pmoles
   6.7 pmoles will yield 1M Ag per bead, which is sufficient for enrichment. If antigen is readily available, 33 pmoles of antigen should be added.
2. Incubate at 4° for >1h.
3. Wash beads: add 1 mL PBSA to beads; place on magnet for 2-5 min.; remove 'supernatant'.
   Repeat.

Cell Selection

1. Measure cell density (OD = 1 corresponds to $1x10^7$ cells/mL).
2. Pellet and wash $375x10^7$ cells. Make two aliquots.
3. Combine cells with 10 uL of bare beads in each tube.
4. Incubate cells + beads at 4° for >2h.
5. Place cells + beads on magnet and collect unbound cells. ** See note below.*
6. Combine cells with 10 uL of new bare beads in each tube.
7. Incubate cells + beads at 4° for >2h.
8. Place cells + beads on magnet and collect unbound cells. ** See note below.*
9. Transfer unbound cells to tubes with washed Ag:beads.
10. Incubate cells + beads at 4° for >2h.
11. Place cells + beads on magnet and remove unbound cells.
12. Wash once with PBSA.
13. Place cells + beads on magnet and remove unbound cells.
14. Resuspend cells + beads in 5 mL SD-CAA.
15. Add 5 uL of cells to 995 uL PBSA (200x dilution).
16. Add 10 uL of dilution to 190 uL PBSA (4,000x dilution).
17. Plate 20 uL of each dilution on SD-CAA plates.

** Also wash, grow, and plate negative selection beads (as in steps 12-17) for comparison.*
(1 col/20 uL)(200x dilution)($5x10^3$ uL in test tube) = $5x10^4$ cells collected per colony
(1 col/20 uL)(40,00x dilution)($5x10^3$ uL in test tube) = $1x10^6$ cells collected per colony

Cell Growth

1. Incubate cells at 30°, 250 rpm for >16h.
2. Remove beads using magnet.
3. Pellet >20x diversity of cells.
4. Resuspend in SG-CAA.
5. Incubate at 30°, 250 rpm for 8-24h to induce protein expression.

Magnetic Bead Sort – Moderate cell number

The current population will be depleted of streptavidin:bead binders and enriched for antigen binders. At least 20x the population diversity should be sorted.

Protein Display

1. Grow and induce yeast population to display Fn3.
2. Use immediately or store yeast at 4°.

Bead Preparation

1. Combine 100 uL of PBSA + 6.7-33 pmoles of biotin-antigen + 10 uL of beads.
   (10 uL beads)(4x10^5 beads/uL)(1x10^6 Ag/bead)(pmol / 6x10^{11} Ag) = 6.7 pmoles
   6.7 pmoles will yield 1M Ag per bead, which is sufficient for enrichment. If antigen is readily available, 33 pmoles of antigen should be added.
2. Incubate at 4° for >1h.
3. Wash beads: add 1 mL PBSA to beads; place on magnet for 2-5 min.; remove 'supernatant'. Repeat.

Cell Selection

1. Measure cell density (OD = 1 corresponds to 1x10^7 cells/mL).
2. Pellet and wash at least 20x library diversity.
3. Combine cells with 10 uL of bare beads.
4. Incubate cells + beads at 4° for >2h.
5. Place cells + beads on magnet and collect unbound cells. ** See note below.*
6. Combine cells with 10 uL of new bare beads.
7. Incubate cells + beads at 4° for >2h.
8. Place cells + beads on magnet and collect unbound cells. ** See note below.*
9. Transfer unbound cells to tubes with washed Ag:beads.
10. Incubate cells + beads at 4° for >2h.
11. Place cells + beads on magnet and remove unbound cells.
12. Wash once with PBSA.
13. Place cells + beads on magnet and remove unbound cells.
14. Resuspend cells + beads in 5 mL SD-CAA.
15. Add 5 uL of cells to 995 uL PBSA (200x dilution).
16. Add 10 uL of dilution to 190 uL PBSA (4,000x dilution).
17. Plate 20 uL of each dilution on SD-CAA plates.

** Also wash, grow, and plate negative selection beads (as in steps 12-17) for comparison.*
(1 col/20 uL)(200x dilution)(5x10^3 uL in test tube) = 5x10^4 cells collected per colony
(1 col/20 uL)(40,00x dilution)(5x10^3 uL in test tube) = 1x10^6 cells collected per colony

Cell Growth

1. Incubate cells at 30°, 250 rpm for >16h.
2. Remove and save beads.
3. Pellet >20x diversity of cells.
4. Resuspend in SG-CAA.
5. Incubate at 30°, 250 rpm for 8-24h to induce protein expression.

C-myc$^+$ FACSort

The population will be sorted for full-length Fn3 clones by selecting clones that contain the c-myc epitope (which would be lost in a truncation or frameshift mutant). The N-terminal epitope HA will also be analyzed to differentitate between plasmid loss and truncation/frameshift.

1. Pellet at least 20x library diversity. Wash with 1 mL PBSA.
2. Resuspend in 50 uL PBSA + 0.25 uL 16B12 (mouseαHA) + 0.5 uL Chαc-myc.
3. Incubate at 22° for >20 min.
4. Wash with 1 mL PBSA.
5. Resuspend in 50 uL PBSA + 0.5 uL GαCh-A488 + 0.5 uL GαM-APC.
   Note that other fluorophore combinations may be used.
6. Incubate at 4° for >15 min.
7. Wash with 1 mL PBSA.
8. Analyze and sort on MoFlo or FACSAria. Collect A488$^+$/APC$^+$ cells (full-length clones).
9. Grow collected cells in ~5 mL SD-CAA.

Zymoprep (Kit II)

Recover plasmids from yeast cells.

1. Measure cell density.
2. Add 10x10^7 cells to a microcentrifuge tube.
3. Centrifuge at 300*g* for 1 min. Remove supernatant.
4. Add 200 uL Solution 1 to pellet.
5. Add 3 uL Zymolyase (take care to use enzyme and not storage buffer).
6. Resuspend by vortexing mildly or pipetting.
7. Incubate at 37° for 15-60 min.

15 minutes will suffice for cells in logarithmic growth. Use >30 min. for cells in stationary phase

8. Add 200 uL Solution 2. Mix gently.
9. Add 400 uL Solution 3. Mix gently.
10. Centrifuge at 12krpm for 8 min.
11. Centrifuge supernatant (in new tube) at 12 krpm for 2 min.
12. Transfer supernatant to Epoch or Qiagen column (not Zymo column).
13. Centrifuge column at 12 krpm for 60s. Discard flowthrough.
14. Add 550 uL of Qiagen buffer PE or Epoch WS. Centrifuge column at 12 krpm for 60s. Discard flowthrough.
15. Centrifuge column at 12 krpm for 60s. Discard flowthrough.
16. Place column in new vial. Add 50 uL elution buffer and let sit for 60s.
17. Centrifuge at 12 krpm for 60s.

Exp. yield: ~5 p/cellx10^8 cells = 5x10^8 p/40 uL=1.25x10^7 p/uL. Use 8 uL in each epPCR

Error-prone PCR

The plasmids for the selected clones will be used as templates for error-prone PCR with nucleotide analogs. Two mutagenesis approaches will be used: mutation of the entire GENE and individual mutation of each LOOP.

1. Prepare PCR mix:

   'Gene' Reaction (1)

5 uL	10x *ThermoPol* buffer	1x *ThermoPol* buffer
2.5 uL	5' primer, 10 uM	0.5 uM 5' primer
2.5 uL	3' primer, 10 uM	0.5 uM 3' primer
1 uL	dNTPs, 10 mM each	200 uM each dNTP
8 uL	zymoprepped DNA,	10^8 plasmids
10 uL	8-oxo-dGTP + dPTP, **20 uM**	**2 uM** each 8-oxo-dGTP, dPTP
20.5	ddH$_2$O	-
0.5 uL	*Taq* DNA polymerase	2.5 units

   'Loop' Reactions (3 total)

5 uL	10x *ThermoPol* buffer	1x *ThermoPol* buffer
2.5 uL	5' primer, 10 uM	0.5 uM 5' primer
2.5 uL	3' primer, 10 uM	0.5 uM 3' primer
1 uL	dNTPs, 10 mM each	200 uM each dNTP
8 uL	zymoprepped DNA,	10^8 plasmids
10 uL	8-oxo-dGTP + dPTP, **200 uM**	**20 uM** each 8-oxo-dGTP, dPTP
20.5	ddH$_2$O	-
0.5 uL	*Taq* DNA polymerase	2.5 units

Sample	5' Primer	3' Primer
gene	W5	W3
BC loop	BC5new	G4bc3
DE loop	G4de5	Lde3
FG loop	G4fg5	FG3new

2. Thermally cycle:

94° for 3 min.	1 step
94° for 45s, 60° for 30s, 72° for 90s.	15 cycles
72° for 10 min.	1 step

Electrophoresis

Purify PCR product by agarose gel electrophoresis. The aim of this step is to remove template DNA and, if possible, any improper PCR product.

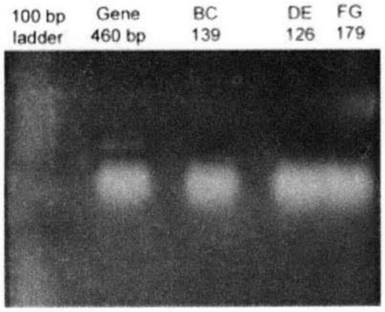
1. Make 1.5% agarose gel (0.75 g agarose, 49 mL ddH_2O, 1 mL 50x TAE).
2. Add 5.6 uL 10x restriction stop buffer to each epPCR tube.
3. Run epPCR samples and ladder at 100V for ~45 min.
4. Dye with SYBR Gold (10 uL in 50 mL 1x TAE) for 20 min.
5. Excise PCR products.

    Gene band should be visible. Loop bands will not be visible.

    Cut a large piece to ensure PCR product recovery.

    ** primer dimers and loop PCR product not discernable **

    OK - step is designed to eliminate template not primer



6. Purify PCR product using Qiaquick Gel Extraction kit. Elute in 40 uL elution buffer.

Amplification PCR

Amplify PCR products to create large quantities of DNA for transformation into yeast.

1. Prepare 200 uL for each epPCR sample.

20 uL	10x *ThermoPol* buffer	1x *ThermoPol* buffer
20 uL	5' primer, 10 uM	1 uM 5' primer
20 uL	3' primer, 10 uM	1 uM 3' primer
4 uL	dNTPs, 10 mM each	200 uM each dNTP
8 uL	extracted PCR product	-
126 uL	ddH_2O	-
2 uL	*Taq* DNA polymerase	-

2. Split into two 100 uL aliquots (for each epPCR sample).
3. Thermally cycle:

94° for 3 min.	1 step
94° for 45s, 60° for 30s, 72° for 90s.	30 cycles
72° for 10 min.	1 step

Sample	5' Primer	3' Primer
gene	GeneAmp5	GeneAmp3
BC loop	BCamp5	BCamp3
DE loop	DEamp5	DEamp3
FG loop	FGamp5	FGamp3

These primers are shorter versions of the mutagenic primers; the oligos are shorter.

Concentrate

Concentrate amplification PCR products for use in transformation.

1. Combine two identical *ampPCR* GENE samples into a 1.7 mL tube.
2. Combine six *ampPCR* LOOP samples into in a 2 mL tube.
3. Add 2 uL PelletPaint to each tube.
4. Add 20 and 60 uL 3M NaAc. Mix briefly.
5. Add 500 and 1200 uL 100% ethanol. Incubate at room temperature for 2 min.
6. Centrifuge at 12krpm for 5 min. Remove supernatant.
7. Add 500 uL 70% ethanol. Vortex briefly.
8. Centrifuge at 12krpm for 5 min. Remove supernatant.
9. Add 500 uL 100% ethanol. Vortex briefly.
10. Centrifuge at 12krpm for 5 min. Remove supernatant.
11. Air dry.
12. Add 1 uL of ddH_2O to pellet once dry.

Electroporation Transformation

1. Grow EBY100 in YPD overnight.
2. Inoculate 50 mL of YPD with 5×10^7 cells (OD = 0.1).
3. Incubate culture at 30°, 250 rpm until cell density is $1.3\text{-}1.5 \times 10^7$ cells/mL.
   This will generally take 6-8 hours. Note that the culture can be inoculated at a lower density (in step 2) and incubated for a longer time, e.g., *overnight.*

Prepare Competent Cells

Chill electroporation cuvettes and buffer E

1. Pellet cells (3,000 rpm in large centrifuge for 3 min.). Remove supernatant.
2. Resuspend in 25 mL of 100 mM lithium acetate, 10 mM Tris, pH 7.5, 1 mM EDTA.
3. Incubate at 30° for 15-60 min.
4. Add 0.096 g DTT to 0.5 mL 1M Tris, pH 8.0. Filter sterilize into EBY100.
5. Incubate at 30°, 250 rpm for 15 min.

from now until after electroporation, keep cells chilled on wet ice and use chilled rotors for centrifugation

6. Centrifuge cells at 2500*g*, 4° for 3 min. Discard supernatant.
7. Wash with 25 mL of buffer E (10 mM Tris, pH 7.5, 270 mM sucrose, 1 mM $MgCl_2$).
8. Centrifuge cells at 2500*g*, 4° for 3 min. Discard supernatant.
9. Resuspend in 1 mL buffer E. Transfer to 1.5 mL tube.
10. Centrifuge at 5000*g*, 4° for 1 min. Discard supernatant.
11. Resuspend in 1 mL buffer E.
12. Centrifuge at 5000*g*, 4° for 1 min. Discard supernatant.
13. Resuspend cells to a total volume of 300 uL.

Electroporation

1. Prepare samples in a microcentrifuge tube: 150 uL cells + 2 uL vector (4 ug) + all insert .
   vector: pCT-Fn3-Gene or pCT-Fn3-Loop (digested at 3 sites)
2. Transfer mixture to two electroporation cuvettes and incubate on ice for about 5 min.
3. Pulse at 25 uF, 0.54 kV (for 2 mm cuvette). Time constant should be 15-45 ms.
4. Add 1-2 mL room temperature YPD to cuvette. Transfer to 14 mL Falcon tube.
5. Incubate at 30°, 250 rpm for 1-2 hours.
6. Centrifuge cells at 1300*g* for 1 min.
7. Resuspend in 1 mL SD-CAA. Transfer to flask with 50-1000 mL SD-CAA.
8. Plate serial dilutions on SD-CAA plates to determine transformation efficiency.
   typical: $\sim 10^6\text{-}10^7$ transformants, but lower efficiencies sometimes occur
9. Incubate culture at 30°, 250 rpm.
10. Incubate plates at 30°.

Fn3 Production in Bacteria

Transformation

1. Thaw Rosetta(DE3) cells [Novagen 70954] on wet ice.
2. Add 20 uL of cells + 0.5 uL of pETh(k)-Fn3 plasmid miniprep to a microfuge tube.
3. Incubate on ice for 5 min.
4. Heat shock at 42° for 30s.
5. Incubate on ice for 2 min.
6. Add 80 uL of room temperature SOC.
7. Incubate at 37°, 250 rpm for ~1h.
8. Plate 20 uL on LB+kan+chlor plate.

Grow Starter Culture

1. Add ~5 mL of LB+kan+chlor to a test tube.
2. Transfer a single colony of Rosetta + pETh(k)-Fn3 to the test tube.
3. Incubate culture at 37°, 250 rpm for 10-24h.

Grow and Induce Large Culture

1. Add 100 [1000] mL of LB (no antibiotics) to a 0.25 [2] L flask.
2. Add 0.1 [1] mL of saturated culture to the flask.
3. Incubate at 37°, 250 rpm for until A_{600} ~1.0 (0.1-1.5 is ok).
4. Add 0.1 [1] mL of 500 mM IPTG to yield 0.5 mM IPTG.
5. Incubate at 37°, 250 rpm for 3-24h.

Prepare Lysate

1. Pellet cells (*e.g.* 2500*g* for 15 min.). Remove supernatant.
2. Resuspend in 5 [50] mL of lysis buffer (50 mM phosphates, pH 8.0, 0.5M NaCl, 5% glycerol, 5 mM CHAPS, 25 mM imidazole, 1x protease inhibitors).
3. Freeze cells at -70°.
4. Thaw cells in room temperature water bath.
5. Lyse cells by sonication or 3x freeze/thaw.
6. Centrifuge lysate at 15,000*g* for 30 min.
7. Save supernatant. Filter if necessary.

Protein Purification

1. Purify Fn3 from lysate using TALON resin.

Benjamin J. Hackel

885 River Meadows Drive, Sheboygan Falls, WI 53085
617-894-2161 benhackel@gmail.com

EDUCATION

Massachusetts Institute of Technology Cambridge, MA
Ph.D. Candidate, Chemical Engineering 2004 - Present
- GPA: 5.0 / 5.0
- Minor: Biological Chemistry

University of Wisconsin Madison, WI
B.S., Chemical Engineering 1999 - 2003
- GPA: 3.99
- Elective Courses: Biochemistry, Biology, Cell Biology, Bioinstrumentation, Biomaterials

EXPERIENCE

Massachusetts Institute of Technology Cambridge, MA
Graduate Research Assistant 2005 - Present
- Advisor: Prof. K. Dane Wittrup
- Thesis: Fibronectin Domain Engineering

University of Wisconsin Madison, WI
Undergraduate Research Assistant 2002 - 2004
- Advisor: Asst. Prof. Eric V. Shusta
- Conducted research towards the development of non-invasive targeted neurotherapeutics
- Developed yeast expression systems to optimize protein production via genetic and cellular engineering

Lawrence Livermore National Laboratory Livermore, CA
Technical Scholar / Researcher Summer 2003
- Conducted research on synthetic peptide chemistry and its applications
- Developed chemistry to synthesize C-terminal α-thioester peptides for use in native chemical ligation of large proteins

Procter & Gamble Cincinnati, OH
Product Development Engineer – Intern Summer 2001
- Conducted laboratory research in upstream chemical product development
- Designed a chemical formulation modification to introduce novel properties into a proprietary surfactant product

TEACHING / SERVICE

University of Wisconsin Madison, WI
- Tutor: Chemical Engineering

Massachusetts Institute of Technology Cambridge, MA
- New England Structural Biology Association Conference Co-organizer
- Teaching Assistant: Integrated Chemical Engineering
- Research Mentor: Undergraduate Research Opportunities Program

Atul Kapila	Biology, MIT	*Fibronectin Production and Stability*
Selasie Goka	Biochemistry, Wellesley	*Fibronectin Crystallization*
Minah Shahbaz	Chemical Engineering, MIT	*Synthetic Library Design*
Danielle Wang	Chemical Engineering, MIT	*Synthetic Library Design*

PUBLICATIONS

Hackel, B.J., Neil, J., White, F., and Wittrup, K.D., “Epidermal growth factor receptor downregulation with heterobivalent fibronectin constructs” *in preparation.*

Hackel, B.J., Ackerman, M., Howland, S., and Wittrup, K.D., “Design, construction, and validation of an improved fibronectin domain library through tailored diversity and stability conservation” *in preparation*

Hackel, B.J., and Wittrup, K.D., “The full amino acid repertoire is superior to serine/tyrosine for selection of high affinity immunoglobulin G binders from the fibronectin scaffold” *submitted.*

Ackerman, M., Levary, D., Tobon, G., Hackel, B.J., Orcutt, K.D., and Wittrup, K.D., “Highly avid magnetic bead capture: an efficient selection method for *de novo* protein engineering utilizing yeast surface display” *Biotech. Prog.* 2009 *25*, 774-783.

Hackel, B.J., and Wittrup, K.D., “Yeast surface display in protein engineering and analysis” in Protein Engineering Handbook (Lutz, S., and Bornscheuer, U.T., eds) 2008, 621-648. Wiley.

Hackel, B.J., Kapila, A., and Wittrup, K.D., “Picomolar affinity fibronectin domains engineered utilizing loop length diversity, recursive mutagenesis, and loop shuffling” *J. Mol. Biol.* 2008 *381*, 1236-1252.

Lipovsek, D., Lippow, S.M., Hackel, B.J., Gregson, M.W., Cheng, P., Kapila, A., and Wittrup, K.D., "Evolution of an interloop disulfide bond in high-affinity antibody mimics based on fibronectin type III domain and selected by yeast-surface display: Molecular convergence with single-domain camelid and shark antibodies" *J. Mol. Biol.* 2007 *368*, 1024-1041.

Chao, G., Lau, W.L., Hackel, B.J., Sazinsky, S.L., Lippow, S.M., and Wittrup, K.D., "Isolating and engineering human antibodies using yeast surface display" *Nature Protocols* 2006 *1*, 755-768.

Hackel, B.J., Huang, D., Bubolz, J.C., Wang, X.X., and Shusta, E.V., "Production of soluble and active transferrin receptor-targeting single-chain antibody using *Saccharomyces cerevisiae*" *Pharm. Res.* 2006 *23*, 790-797.

Camarero J.A., Hackel, B.J, DeYoreo, J.J, and Mitchell, A.R., "Fmoc-based synthesis of peptide alpha-thioesters using an aryl hydrazine support" *J. Org. Chem.* 2004 *69*, 4145-51.

PRESENTATIONS

Hackel, B.J.* and Wittrup, K.D., "Synthetic library design" (Poster) International Conference on Biomolecular Engineering, Santa Barbara, CA January 2009.

Hackel, B.J.* and Wittrup, K.D., "EGFR antagonists based on the fibronectin domain" American Institute of Chemical Engineers Annual Meeting, Philadelphia, PA November 2008.

Hackel, B.J.* and Wittrup, K.D., "Synthetic library design" American Institute of Chemical Engineers Annual Meeting, Philadelphia, PA November 2008.

Hackel, B.J.* and Wittrup, K.D., "EGFR targeted Fn3 domains" (Poster) Biomedical Engineering Society Annual Fall Meeting, St. Louis, MO October 2008.

Hackel, B.J.* and Wittrup, K.D., "High affinity Fn3 domains" American Chemical Society National Meeting, Boston, MA, August 2007

Hackel, B.J.*, Kapila, A., and Wittrup, K.D., "Fibronectin engineering" (Poster) Symposium of the Protein Society, Boston, MA July 2007.

Hackel, B.J.* and Wittrup, K.D., "Synthetic library design" (Poster) PEGS: Protein Engineering Summit, Boston, MA May 2007.

Hackel, B.J.*, Kapila, A., and Wittrup, K.D., "High affinity Fn3 engineering" (Poster) International Conference on Biomolecular Engineering, Coronado, CA January 2007.

Hackel, B.J. and Wittrup, K.D*., "3 pM Fn3 HEL binders from only 10^8 clones" Protein Engineering for Biotherapeutics Conference, San Diego, CA December 2006.

Hackel, B.J.* and Wittrup, K.D., "High affinity Fn3 domains", American Institute of Chemical Engineers Annual Meeting, San Francisco, CA November 2006.

Hackel, B.J. and Shusta, E.V.*, "Yeast expression of antibody-targeted neurotrophins" 227th American Chemical Society National Meeting, Anaheim, CA March 28, 2004.

Hackel, B.J., Wang, X.X., and Shusta, E.V.*, "Expression of brain-targeted therapeutic proteins in yeast", American Institute of Chemical Engineers 2003 Annual Meeting, San Francisco, CA November 16-21, 2003.

HONORS

Graduate
- National Defense Science and Engineering Graduate Fellowship
- National Science Foundation Graduate Fellowship
- International Conference on Biomolecular Engineering Poster Award

Undergraduate
- Hilldale Undergraduate Research Fellow
- Wisconsin Academic Excellence Scholar
- Wisconsin All-State Scholar
- Herbert V. Kohler Award
- C.S. Brown Chemical Engineering Award
- R. and E. Pott Engineering Award
- M.A. Crosby Chemical Engineering Award
- Phi Eta Sigma National Scholar