Title

mRNA Display selection of a novel activated leukocyte cell adhesion molecule (ALCAM) binding protein from a modified combinatorial protein library based on the tenth domain of human fibronection III (10FnIII)

Author

Park, Ann

Publication Date

2015

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

mRNA Display selection of a novel activated leukocyte cell adhesion molecule (ALCAM) binding protein from a modified combinatorial protein library based on the tenth domain of human fibronection III (10FnIII)

A thesis in partial satisfaction of the

requirements for the degree Master of Science

in Physiological Sciences

by

Ann Nahhyun Park

2015

ABSTRACT OF THE THESIS

mRNA Display selection of an Activated Leukocyte Cell Adhesion Molecule (ALCAM) binding protein from a modified protein library based on the tenth domain of human fibronection III (10FnIII)

by

Ann Nahhyun Park

Master of Science in Physiological Science

University of California, Los Angeles, 2015

Professor Christopher T. Denny, Co-Chair

Professor David Glanzman, Co-Chair

Antibodies have served as the preeminent model for proteins tailored to exhibit specific binding properties. Unfortunately, the development and manufacturing of antibodies is an expensive and time-consuming process. Recently, however, the field of drug targeting has gravitated towards developing smaller, alternative binding proteins that harness the targeting power of antibodies,

but are produced at more expedient rates and with lower costs. *In vitro* display platforms allow for the discovery of novel binding proteins from highly complex libraries of non-immunoglobulin scaffolds. The tenth domain of human fibronectin III (10FnIII) molecule is one of these extensively studied antibody mimics that demonstrate promise as an antibody alternative in targeted therapeutics and in diagnostic imaging. In this study, mRNA display was applied to screen for a novel binding protein specific to a known cancer biomarker, activated leukocyte cell adhesion molecule (ALCAM), in a previously modified combinatorial protein library based on the tenth domain of human fibronectin III (e10FnIII). Iterative rounds of affinity selection resulted in the discovery of a single e10FnIII variant, designated Fn16.3, which specifically binds ALCAM *in vitro*. In bacteria, Fn16.3 was robustly expressed but formed insoluble aggregates. These results demonstrate that *in vitro* selection can be used to isolate novel binding proteins, but that further evolution of functional clones may be required to generate a binding molecule that can also be expressed in a desirable expression system.

The thesis of Ann Nahhyun Park is approved.

Rachelle Crosbie-Watson

David Glanzman, Committee Co-Chair

Christopher T. Denny, Committee Co-Chair

University of California, Los Angeles

2015

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGMENTS

This thesis represents the culmination of a long, arduous and unexpected journey. During this project I had to work through both academic and personal challenges and I am grateful to the individuals below for helping me become a better scientist and person as a whole.

I would like to thank two of my committee members, **Dr. David Glanzman** and **Dr. Rachelle Crosbie-Watson** for taking the time to review this thesis work.

To the former members of the Denny Lab- **Kelly, Jenny, Noah, Jason, Neha** and **Catherine**- thank you for commiserating with me through failed experiments and lab crises, and for our many fun and funny memories. **Jenny**, you pulled me back into the fold when I lost focus and I couldn't be where I am now without your empathy and encouragement. Thank you for your friendship.

To my family- thank you for always wanting the best for me and always pushing me to do my best in everything. **Bryan Park**- you are my rock, my best friend, my better half.

I would like to thank **Dr. Anders Olson**, whose guidance made this project possible. Thank you counseling me through library construction and mRNA display protocols and for giving me advice when I hit (many) roadblocks along the way.

Most importantly, I would like to extend my gratitude to **Dr. Christopher Denny**, my thesis (and life) advisor since 2008. Thank you for unconditionally supporting my academic and personal goals and for standing by me in my darkest times. You're not just my mentor, but also my role model and father away from home.

INTRODUCTION

Over the last century, antibodies have served as the archetype of binding proteins. Immunoglobulins are incredibly versatile molecules that can be generated against a wide range of antigens. Physiologically, they play a crucial role in humoral immunity by identifying and clearing foreign antigens. Moreover, taken out of their natural context, antibodies can be generated against virtually any antigen and therefore can be employed in a boundless number of research and medical applications. Immunoassays are central to many laboratory experiments *e.g.*, immunohistochemistry, Western blot and flow cytometry. Clinically, immunoassays have traditionally served as diagnostic tools in such contexts as serological measurements of viral or microbial antigens, and blood typing. Immunoassays can also be manufactured for the general populace *e.g.*, over the counter pregnancy tests that measure levels of human chorionic gonadotropin (HCG).

In recent decades, more sophisticated progress in antibody engineering has made it possible for monoclonal antibodies to be administered as therapies themselves. As of 2009, over 20 monoclonal antibodies were available for therapeutic purposes in the United States and European Union (1). A few well-known antibody-based therapies include Adalimumab (trade name Humira), which inhibits TNF-α signaling in inflammatory autoimmune diseases such as rheumathoid arthritis and Crohn's disease; Bevacizumab (trade name Avastin), which blocks vascular endothelial growth factor A (VEGF-A) mediated angiogenesis in cancer; and Trastuzumab (trade name Herceptin), which blocks HER2/neu receptor signaling in certain breast cancers.

However, inherent disadvantages to using antibodies as therapeutic reagents have come to light. Monoclonal antibodies are physically bulky ~150 kDa molecules comprised of four polypeptides linked by disulfide bridges. Accordingly, because of the structural complexity of antibodies, the use of a eukaryotic expression system is the only viable option in antibody production. Eukaryotic hosts like Chinese hamster ovary cells and hybridoma cell lines harbor cellular machinery that can modify nascent proteins post-translationally and provide chaperones that facilitate proper folding. Unfortunately, antibody production by eukaryotic expression systems is an expensive process, as they require large quantities of cells that grow slowly and produce low levels of protein. Moreover, antibody manufacturing often requires an extensive and optimized purification process that compounds the costs of commercializing antibodies. Bacterial hosts are the preferred mode of protein production because they grow relatively quickly (with a doubling time of 20 minutes versus ~24 hours for typical eukaryotic cells), express much higher levels of protein, and can be easily scaled up for large format production.

The large size of antibodies also imposes a limitation on their pharmacokinetic efficacy as targeting molecules, particularly with respect to their ability to extravasate through vasculature and penetrate tissues in transit to target cells. Solid tumors, which represent a majority of human cancers, often harbor complex and extensive vasculature networks and high interstitial pressure, which makes antibody diffusion to target tissues considerably more difficult (1).

The field of antibody engineering has attempted to address antibodies' size problem by progressively pruning away domains not required for antigen recognition. Relative to the remaining portions of the molecule, the binding functionality of antibodies is confined to a small region called the complementarity determining region (CDR) which is formed by the variable

heavy (V_H) and light (V_L) chain domains. Single chain variable fragments (scFv), which are synthesized by joining together V_H and V_L chains by a flexible peptide linker, make up the smallest portion of antibodies that retain their original binding ability. scFv's can also be fused together to create multivalent diabodies, triabodies and tetrabodies (2).

Recently, the use of non-immunoglobulin molecules as antigen-targeting agents, dubbed "antibody mimetics," has gained traction in biomedical research. Protein scaffolds can be engineered to circumvent the structural limitations and high production costs associated with antibodies while sustaining their binding capacities. They are often small monomers that are structurally stable and are easily expressed at high levels in microbial systems. Moreover, engineered scaffolds can achieve higher binding affinities (pM) than typical antibodies (μM-nM). As a result, these protein scaffolds are highly attractive alternatives to antibodies. A few examples of well-studied scaffolds are designed ankyrin repeat domains (DARPins), affibodies modeled after the *staphylococcal* Z-domain of protein A, and monobodies based on the tenth domain of human fibronectin III (3).

In protein engineering, directed evolution is a powerful tool to improve specific protein properties or to develop novel functions in proteins. Directed evolution relies on the Darwinian principle of natural selection that enables the survival of favorable mutants when faced with selective pressures. A starting protein can be genetically diversified by techniques like error prone PCR, chemical mutagenesis or gene shuffling to create a large pool of variant molecules. This diverse pool is then subjected to selection for a desired trait, where only the most "fit" variants are enriched. One major advantage of employing a directed evolution strategy is that prior knowledge of protein structure is not required, and oftentimes the strategy results in the discovery of functional variants with unexpected mutations. Additionally, selection conditions

can be modified to evolve proteins for enhanced binding affinities for ligands (termed affinity maturation) for increased solubility, and for improved temperature or pH specific stability.

Display technologies are becoming important techniques for the directed evolution of proteins with novel functions from combinatorial libraries. Protein function can be selected for without prior knowledge of the encoding sequence. This is achieved by establishing a physical tether between genes and their encoded progeny. In other words, knowledge of the gene sequence is not required until a favorable functionality is achieved.

There are two types of display technologies – *in vivo* and *in vitro* display. Protein selection using *in vivo* display operates by fusing library sequences to a host gene such as a phage coat filament or the yeast Aga2p surface protein. Cell-based display, however, is limited by transformation efficiencies and growth rates of host organisms. For example, phage display is limited to a library complexity of 10^9 different molecules (4). *In vitro* display methods bypass the need for a transformation step and have the capacity to handle libraries of $>10^{13}$ members (5). In comparison, somatic recombination in humans limits antibody specificities to 10^{11}. Enlarging library complexities increases that probability of identifying rare proteins that are likely important for binding highly specific epitopes and/or small ligands.

mRNA display is an *in vitro* display technique that directly links a translated peptide to its parent mRNA via a stable covalent linkage to the antibiotic puromycin (Fig. 1). After run-off *in vitro* transcription of a DNA library, transcripts are ligated to a short oligonucleotide linker that attaches a puromycin molecule to the mRNA's 3'end. Part of puromycin structurally resembles the tyrosyl-tRNA, which enters the ribosome's A-site during translation to form a covalent bond between the nascent peptide and its progenitor transcript. Reverse transcription of

the mRNA molecule stabilizes the nucleic acid molecule and provides template sequences for amplification after affinity enrichment. The fused cDNA/mRNA-protein molecules can then be applied to an immobilized target of interest. Only molecules with high affinity for the target will be bound and the rest can be washed away, resulting in a pool of DNA sequences encoding a desired trait.

mRNA display is a versatile protein selection technique that can be utilized to find proteins binding to other proteins, nucleic acids, or other small molecule ligands. It can also be applied to studying the mechanism of protein-ligand binding, to enhance the binding affinity and/or structural stability of a known protein-ligand interaction.

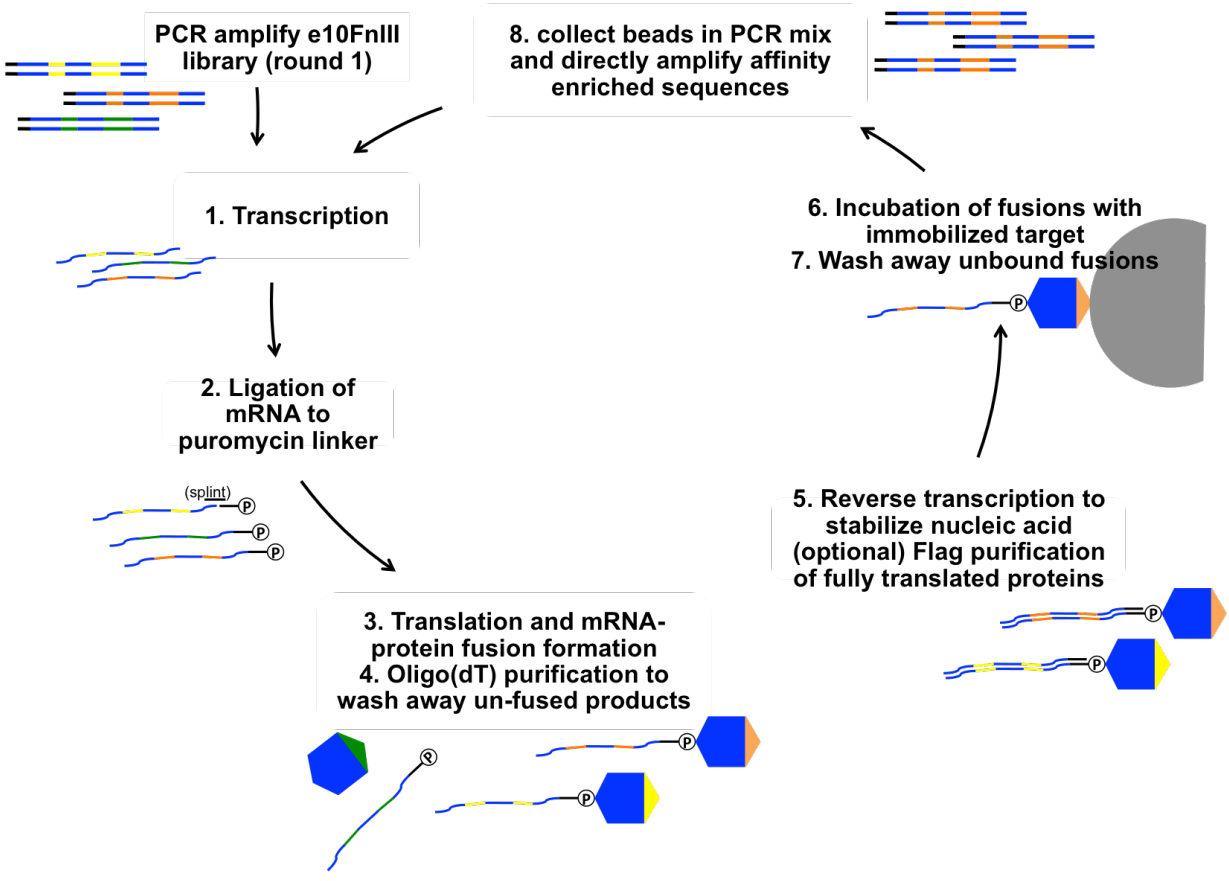

Figure 1. mRNA display scheme. Adopted from Olson et al. (6).

This thesis projected utilized mRNA display selection on a modified combinatorial library based on the tenth domain of human fibronectin III (10FnIII). 10FnIII is a well-characterized scaffold that is a promising alternative to antibodies as a targeting molecule. Activated Leukocyte Cell Adhesion Molecule (ALCAM) was chosen as the target ligand for mRNA display. As a known cancer biomarker, ALCAM is a relevant candidate for targeted therapy and diagnostic applications.

10FnIII is an evolutionary conserved protein that is found across a wide array of organisms. Structurally, it is homologous to the immunoglobulin VH domain and is comprised of a beta-sandwich with three exposed loops- BC, DE, and FG- that correspond to the VH domain's CDR1, CDR2, and CDR3 regions. The BC, DE and FG loops are comprised of 7, 4 and 10 residues, respectively. 10FnIII is also one of the oldest and widely studied scaffolds in protein engineering. 10FnIII is a small 94 amino acid monomer without disulfide bonds or glycosylated moieties, which makes it amenable to bacterial expression. Because it is derived from a naturally occurring human fibronectin, it risks lower immunogenic response than molecules derived from animal systems. Its compact and stable structure also allows for chemical modifications like pegylation and the addition of reactive groups like cysteine for conjugation to therapeutic agents.

Fibronectin derived scaffolds called Adnectins are already in clinical use. CT-322 is a pegylated Adnectin that blocks vascular endothelial growth factor receptor-2 (VEGFR-2) signaling and VEGFR-2 mediated primary tumor angiogenesis, and is currently in phase II clinical trials for the treatment of glioblastoma (7,8). Adnectins have also been fused to produce bi-specific molecules, coined El-Tandems, that have been able to simultaneously block epidermal growth factor receptor (EGFR) and insulin-like growth factor-I receptor (IGF-IR) signaling that drives tumorigenesis in several human cancer cell lines (9).

In 1998, Koide et al. reported the first instance in which a phage display library of 10FnIII was prepared by randomizing BC and FG loops regions, and which was used to select for high affinity ubiquitin binders (10). Xu et al. took diversification a step further and randomized all three (BC, DE, FG) loops to construct an mRNA library of $>10^{12}$ complexity that selected for TNF-α binders with dissociation constants (K_d) 1-24 nM (11).

The e10FnIII library used in this thesis study was first developed by Olson and Roberts (12) and differed from the Koide et al. and Xu et al. libraries in several respects (Fig. 2). Only the BC and FG loops were randomized, and seven unstructured N-terminal residues were eliminated. The scaffold backbone also was modified with five solubilizing mutations indicated by asterisks in Fig. 2 (6). Selected e10FnIII variants have been shown to target modification specific phospho-IκBa, attenuate SARS replication by intracellularly targeting viral nucleocapsid proteins, and inhibit IL-6 signaling in human hepatocytes (6,13,14).

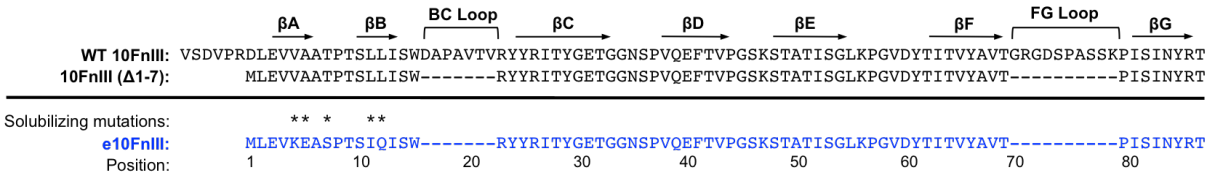

Figure 2. 10FnIII sequence comparisons. Wild-type (WT) 10FnIII is aligned against the previously described 10FnIII library with randomized BC and FG regions and lacking seven unstructured amino-terminal residues. The modified e10FnIII library with 5 solubilizing mutations is highlighted in blue (6,12).

The target molecule used in this thesis project ALCAM, is also known as CD166. ALCAM belongs to the immunoglobulin superfamily (IgSF), a large family of proteins that share

the Ig-fold, a sandwich structure comprised of beta-sheets. Ig domains are categorized as either variable (V-type) or constant (C-type). ALCAM is a ~110 kDa transmembrane glycoprotein that consists of two V-type and three C-type domains (Fig. 3). First identified as a ligand to the CD6 membrane protein expressed by activated leukocytes, ALCAM is involved in both homophilic and heterophilic cell-cell interactions that regulate cell growth and migration in different tissues (15,16). The amino-terminal V1 domain is essential for CD6 ligand binding and homophilic ALCAM-ALCAM cell interactions (17). The three membrane proximal Ig C-type domains mediate homotypic oligomerization at the cell surface (18). Alternative splicing also gives rise to a soluble ALCAM isoform consisting of only the V1 domain, which has been shown to modulate endothelial cell growth and migration, and influence ALCAM-ALCAM interactions (19). ALCAM is overexpressed in a multitude of cancers that include glioblastoma (20), pancreatic cancer (21), colorectal cancer (22), and head and neck squamous cell carcinoma (23).

As a widely identified cancer biomarker, ALCAM has the potential to be exploited as a target for applications such as tissue imaging and drug delivery. In fact, an anti-ALCAM cys-diabody has already been generated for the purposes of imaging ALCAM positive tissues via positron emission tomography (PET) (24). The same anti-ALCAM cys-diabody has also been conjugated to liposomal nanoparticles designed to deliver doxorubicin, a cytotoxic chemotherapeutic drug, to osteosarcoma cells (25).

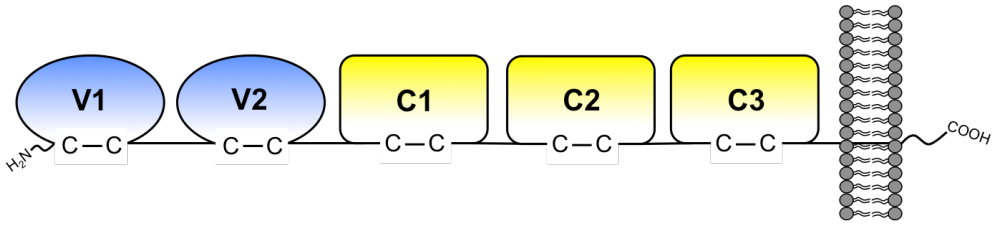

Figure 3. Schematic of ALCAM structural domains. The distal portion of ALCAM is made up of two V-type domains followed by three membrane proximal C-type domains and a relatively short cytoplasmic tail.

The purpose of this study was to find a non-antibody molecule that can bind specifically to a medically relevant cancer biomarker, ALCAM. mRNA display was used to select for anti-ALCAM binding protein from an e10FnIII combinatorial protein library which, as described, overcomes the structural limitations to antibodies in medically relevant applications. Findings show that mRNA display was able to select for a single e10FnIII variant designated as Fn16.3 that specifically recognizes both full length ALCAM and a truncated version consisting only of the V1-V2 domains (vALCAM). Expression experiments, however, revealed that Fn16.3 has limited solubility when produced in a bacterial expression system. Additional experiments to find another ALCAM binder were unsuccessful, as they ultimately pointed to Fn16.3 as the sole binder. As a result, further evolution to enhance soluble expression is required if Fn16.3 is to be a practical alternative to an anti-ALCAM antibody.

MATERIALS AND METHODS

Oligos used in e10FnIII library construction and pAO5, pAO9 and pJD1 were generously provided by Dr. C. Anders Olson from the Department of Molecular and Medical Pharmacology, at UCLA. Illumina sequencing reagents and batch processing of HTS data was graciously provided by Dr. C. Anders Olson and Nicholas C. Wu from the Department of Molecular Pharmacology at UCLA.

Cell culture

Cell lines 293T and KHOS 240S were seeded in Dulbecco's Modified Eagle Medium (HyClone # SH30022.01) supplemented with 10% fetal bovine serum (Gemini Bioproducts). Cell lines were incubated in 5% CO_2 at 37°C.

ALCAM Cloning and Expression

ALCAM sequence was isolated from KHOS 240S, an osteosarcoma cell line that natively expresses ALCAM. Total RNA was extracted from cells (Qiagen RNeasy Kit) and first strand cDNA was generated with oligo(dT) and random hexamers (Invitrogen). ALCAM was amplified from cDNA using primers designed to include the signal peptide, which would allow purification of a soluble product, and to isolate only the two distal V1-V2 variable domains of the protein, termed vALCAM, with primers ALCAM-For and ALCAM-V1V2-Rev. The resultant 797 bp PCR product was TA-ligated into the pCR2.1 cloning vector (Invitrogen). vALCAM was amplified with primers ALCAM-BamHI-For and ALCAM-V1V2-EcoRI-Rev to introduce

BamHI and EcoRI restriction sites for subcloning. BamHI/EcoRI (New England Biolabs) digested products were subcloned into the mammalian expression vector pJD1 (pJD1-vALCAM-KV2.7), which contains the C-terminal 6xHis tag and BirA biotinylation sequence. Ligated products were transformed into chemically competent DH5α *E. coli* and ampicillin selected.

pJD1-KV2.7 was transiently transfected by calcium phosphate into 293T cells. Media from transfected 293T cells was collected at 24, 48 and 72 hours post transfection and stored at -80°C prior to purification. Prior to purification, media was supplemented with protease inhibitor (Roche #11873580001), 500 mM NaCl and 10 mM imidazole. Treated media was affinity purified with nickel-NTA agarose (Qiagen # 30230) and quantified via BCA assay. Purified protein was concentrated with Amicon 30 kDa spin concentrators and quantified via BCA assay. Purified protein was biotin labeled with biotin ligase (GeneCopoeia #BirA500). Expression was verified with SDS-PAGE/coomassie staining and immunoblot. Western blot was performed using anti-ALCAM (R&D Systems #MAB656).

e10FnIII Library assembly

The following described the assembly of an e10FnIII library for vALCAM selection. (An archived library was used for rhALCAM selection). The e10FnIII library was assembled as previously described (6,12,13) using oligonucleotides listed in Table I. Cassettes containing the randomized BC and FG loop regions were synthesized separately then ligated. 16 pmoles of FnOligo3, containing the randomized BC loop, and 16 pmoles of FnOligo4 were annealed and extended with Superscript III® (Invitrogen) in a 65 µl reaction. 40 pmoles of FnOligo7, containing the randomized FG loops, and 40 pmoles of FnOligo6 were annealed and extended with SuperscriptIII® in a 130 µl reaction. Extension for both fragments occurred at 42°C for 1

hour, then column purified (Invitrogen Purelink® PCR Purification Kit). The BC fragment was amplified and extended with FnOligo2 and FnOligo5, which includes a 3' BsaI restriction site, in 400 µl with KOD Hot-Start Polymerase (EMD/Millipore), and column purified (Invitrogen). BC and FG fragments were then digested with BsaI at 50°C for 3 hours, then column purified (Invitrogen). Digested fragments were ligated together in a ~1:1 molar ratio with T4 DNA ligase (New England Biolabs) at room temperature for 1 hour. The ligation reaction was run on a 2% TAE agarose gel and the ~300 bp product was excised and purified (Qiagen QIAquick Gel Extraction Kit). A total of 1.02 pmoles of the ligated product was recovered and amplified in a 1 ml reaction with FnOligo1, which encodes the T7 RNA polymerase promoter region and the tobacco mosaic virus (TMV) translation enhancing sequence, and FnOligo9.2, which encodes the (puromycin) linker region and a flag-tag sequence. KOD Hot-Start polymerase was used for all amplification reactions during mRNA display.

mRNA Display

Library amplification. Round 1 amplification was performed with FnOligo1 and FnOligo9 (rhALCAM selection) or FnOligo9.2 (vALCAM) primers. The forward FnOligo1 primer encodes T7 RNA polymerase promoter and the Tobacco Mosaic Virus (TMV) translation enhancer sequence. The reverse FnOligo9/9.2 primer extends the 3' end of the gene fragment with a flag tag and contains the complementary sequence to the DNA linker. All subsequent selection cycles were amplified with FnOligo1 and FnOligo10 (rhALCAM) or FnOligo10.2 (vALCAM), an abridged portion of FnOligo9/9.2 oligo.

In vitro transcription. Amplified DNA was *in vitro* transcribed with T7 polymerase (Ambion #AM2081) at 37°C for 2 hours or until the formation of a visible precipitate. EDTA was added to a final concentration of 50 mM EDAT to dissolve Mg^{2+} salts. mRNA was purified via column purification (Qiagen RNeasy Kit). RNA concentration was measured via UV spectrophotometry.

Splint ligation of RNA to puromycin linker. 0.8 nmoles of RNA were combined with 1.1 nmoles of DNA splint 1 (rhALCAM) or splint 2 (vALCAM) and 1.2 nmoles of pF30P, and ligated with T4 ligase (New England Biolabs) at for 10 min at 65°C. The ligated product was separated from unligated components by running them out on a 4.25% urea-PAGE gel for 2.5 hrs at 20W. The ligation product was electroeluted at 200 V for 1 hour and ethanol precipitated.

In vitro translation and fusion formation. 40 pmoles were *in vitro* translated in 100 µl of rabbit reticulocyte (Applied Biosystems #AM1200) for 1 hour at room temperature. Fusion formation was facilitated by the addition of KCl (500 mM) and $MgCl_2$ (60 mM) and further incubation for 30 min at room temperature.

Oligo(dT) mRNA purification. Translated products were oligo dT purified in 1 ml reactions with 2 mg oligo dT cellulose (GE Healthcare #27-5543-02) in binding buffer (50 mM Tris, pH 8, 1 M NaCl, 10 mM EDTA, 0.1% Triton X-100). After rotating for 1-2 hours at room temperature, the samples were transferred to 0.22 µm spin filters, and washed with cold TBS with 0.05% Tween-20. The resin was then washed twice with 2.5X Superscript II buffer (Invitrogen). Fusion products were eluted with three 30 µl volumes of warm 5 mM Tris, pH 7.5.

Reverse Transcription. mRNA was reverse transcribed with Fn oligo 10.2 via SuperScriptIII (Invitrogen) at 42°, ~1 hour. The reaction was stopped by adding EDTA to a final concentration of 5 mM.

Flag pre-selection. In order to get rid of untranslated, mis-translated and unfused mRNA, reverse transcribed products were passed through M2 anti-flag beads. 75 µl of beads were washed with binding buffer (500 µl TBS plus 0.05% Tween-20, 5 mM EDTA, 1 mg/ml BSA). After rotating for 1 hour at 4°C, beads were washed four times with TBS plus 0.05% Tween-20, and eluted with 100 µl TBS-T 0.05% supplemented with 0.2 mg/ml 3xFLAG peptide (Sigma #F4799). Flag pre-selection was implemented into selection protocol starting with round 3.

Negative selection. Negative selection was performed twice to remove any variants with nonspecific affinity to streptavidin agarose medium used during affinity enrichment. Prior to affinity enrichment in round 1, reverse transcription products were filtered through 75 µl of empty agarose beads to bind any agarose specific variants. Beads were washed with 500 µl TBS-T 0.05% supplemented with 1 mg/ml BSA and 20 µg/ml free streptavidin to compete with any streptavidin binding elements. In round 5, neutravidin beads were used to immobilize target ALCAM in lieu of streptavidin.

Affinity enrichment. ~2 µg target vALCAM and 10 µl streptavidin beads (Pierce #20357) was combined in 500 µl TBS-T 0.05% in a 0.22 µM Spin-X filter and rotated for 30 minutes at room temperature. Flow thru was removed by centrifugation and beads were washed twice with 500 µl TBS-T 0.05%, then twice with binding buffer. The flag-purified fusions were combined with 400 µl binding buffer and immobilized target, and rotated for 1 hour at room temperature. Beads were washed 4 times with binding buffer and collected in 400 µl KOD Hot-Start polymerase PCR mix (EMD Millipore) made with Fn oligo 1 and 10.2. Amplification was monitored by running 1 µl of the ongoing PCR reaction on a 2% SB gel and stopped when product concentration reached ~80 ng/µl.

PmlI treatment. Archived DNA from previous selection rounds (4 µl or 8% from 50 µl) were amplified for four cycles in 400 µl reactions with Fn oligo 1 and 10.2. After spin column purification (Invitrogen), spectrophotometry measurement indicated yields of 11-13 µg of DNA or ~50 ng/µl. Restriction digests with PmlI (New England Biolabs) were incubated for 2 hrs at 37°C. Digested products were run out on a 2% TAE gel.

Fn 16.3 Cloning and Bacterial Expression

Expression vectors were generously provided by Dr. Anders Olson. Restriction enzymes used during cloning were purchased from New England Biolabs. Fn16.3 was cloned into pAO5 which contains a N-terminal 6x-His tag via XhoI and BamHI restriction sites, or into pAO9, which includes an N-terminal maltose binding protein (MBP) tag, via NdeI and BamHI restriction sites. A cysteine was introduced to the 3' end of Fn via PCR and also subcloned into expression vectors via XhoI and BamHI (pAO5) or NdeI and BamHI (pAO9).

Bacterial expression. Constructs were transformed into chemically competent *E. coli* BL21(DE3) via heat shock at 42°C for 1 min. Inoculated starter cultures were shaken at 250 RPM overnight at 37°C. Cultures were diluted 1:100 and shaken at 37°C until OD_{600} of 0.3-0.5. Expression was induced with 500 mM IPTG for 3 hours. Cells were pelleted and frozen at -20°C prior to purification. Pellets were lysed with BPER lysis buffer (Qiagen #90084) supplemented with protease inhibitor (Roche #04693159001), 300 mM NaCl, and 10 mM imidazole. Lysates were purified via Ni^{+2} NTA affinity chromatography. Lysates were gravity filtered over 0.5-1.0 ml Ni-NTA agarose columns at 4°C. Resin was pre-equilibrated with lysis buffer. Columns were washed with at least three column volumes of wash buffers in increasing imidazole

concentrations; wash buffer 1 consisted of 35 mM Tris, 150 mM NaCl, 10 mM imidazole, 0.5X BPER, 5% glycerol, final pH 7.2, and wash buffer 2 consisted of 50 mM Tris, 300 mM NaCl, 25 mM imidazole pH 7.2, 10% glycerol at final pH 6.8. Purified protein were eluted in 50 mM Tris, 300 mM NaCl, 200 mM imidazole, pH 7.2, 10% glycerol, protease inhibitor, with a final pH of 6.8. Purification samples were analyzed by SDS-PAGE (with 15% gels) followed by Coomassie blue (G-250) staining.

Two-step purifications were performed with MBP fusion. Following 6xHis purification as described above, pooled eluates were loaded into amylose resin packed columns at room temperature. Resin was washed with 10 column volumes TBS. MBP tagged proteins were eluted with 10 mM maltose in TBS.

Fn16.3 Cloning and Mammalian Expression

Fn16.3 was cloned into p3xFLAG-CMV14 for expression in mammalian culture. The insert was amplified from pAO5-Fn16.3 with primers that introduced a new 5' EcoRV restriction site (Fn-EcoRV-5138-F) while maintaining the 3' BamHI recognition site from pAO5 (Fn-BamHI-5140-R).

p3xFLAG-CMV14 was transiently transfected into 293T cells with calcium phosphate. Lysates were harvested 24 hours post-transfection. Cells were twice washed with cold PBS and lysed with NP-40 lysis buffer. Lysates were collected and incubated on ice for 30 minutes before centrifugation. Fn16.3-3xFLAG was purified with anti-Flag M2 affinity gel (Sigma #A2220) and eluted with 3XFLAG peptide (Sigma #F4799).

Cell lysate concentrations were measured with Qubit Fluorometer (Invitrogen). 500 μg lysates were added to 5 μl M2-Flag agarose resin in a 0.22 μM cellulose acetate centrifuge tubes (Costar Spin-X tubes, cat#8160) and final volume brought up to 500 μl with PBS-T 0.1% (PBS, pH 7.4, 0.5% Tween-20) supplemented with protease inhibitor (Roche). Cell extracts were rotated with beads for 1 hour at 4°C. The lysate/bead suspension was centrifuged to remove unbound proteins. Beads were washed three times with 500 μl PBS-T 0.1%. Fn16.3 was eluted by incubating beads with 2x SDS loading dye. Purification samples were run out on a 15% polyacrylamide gel and immunoblotted with M2 anti-Flag (Sigma F1804) diluted 1:1000.

Binding experiments

For ALCAM pull down assay, Fn16.3 was isolated from cell lysates as described above. Instead of adding 2xSDS loading dye after washes, ~2 μg unbiotinylated vALCAM was added to Fn16.3 bound beads in 500 μl PBS-T (0.1%) and rotated for 1 hr at 4°C. Unbound vALCAM was centrifuged out and beads were washed three times with PBS-T 0.1%. Bound complexes were eluted with 2xSDS loading dye and run out on a 15% polyacrylamide gel. Immunoblot was performed with either anti-His (Qiagen #34660) diluted 1:2000 in 3% BSA or anti-ALCAM (R&D Systems #MAB656) diluted 1:2000 in 5% milk.

Illumina sequencing

rhALCAM selection pool 4B and vALCAM selection pools 3 and 5P were modified atBpmI recognition sites were added 5' to BC and FG variable loop coding strands. The recognition

sequences were placed so that the type II restriction enzyme would leave 5 bp buffer regions adjacent to the randomized loops. rhALCAM selection pool 4B and vALCAM selection pools 3 and 5P were amplified to increase variant copy numbers ~100-fold (7 cycles) with BpmI encoding primers, Fn-BpmI-For and Fn-BpmI-Rev (Table 2). 1 μg of each amplified pool was digested with BpmI (New England Biolabs) at 37°C for 1 hour. Digested fragments were gel purified (Qiagen QIAquick Gel Extraction Kit) and end repaired with T4 polymerase (Invitrogen) for 15 min at room temperature to remove 2 bp overhangs left by BpmI. Samples were purified by ethanol precipitation and phenol-chloroform extraction, yielding ~2 pmoles of each pool. 3'-A overhangs were added to sequences by incubation with Taq polymerase (New England Biolabs) and dATP for 20 min at 72°C, then purified by ethanol precipitation. Sequences were then TA-ligated to adapters present in 1000-fold excess with T4 ligase (New England Biolabs). Adapters encoded ACT, ATG and TCG barcodes to identify selection pools from which sequences were obtained during data analysis. TA-ligated sequences were amplified with primers for flow cell annealing. Sequencing was performed via Illumina HiSeq 2000, available at the UCLA Clinical Microarray Core facility.

Cloning and expression of HTS identified e10FnIII clones

e10FnIII variants identified via HTS were generated by site-directed, ligase-independent mutagenesis (SLIM) (26,27) using pAO5-Fn16.3 as a template. Loop sequence specific tailed primers were designed to swap out BC and FG loop regions in the pAO5-Fn16.3 plasmid. Fn69 and Fn61 were labeled based on their copy number enrichment in rhALCAM pool 4B (Fig. 11). SLIM was used to first introduce BC loop sequences.

Two PCR reactions were set up as 25 µl reactions with 0.1 ng/µl template concentration and paired short forward (Fs) and tailed reverse (Rt) primers, or tailed forward (Ft) and short reverse (Rs) primers. A high fidelity KOD Hot-Start polymerase was used to amplify the ~6 kb plasmid (EMD/Millipore). PCR reactions were then incubated with DpnI (New England Biolabs) for 1 hour at 37°C to digest methylated template plasmid. Tailed PCR products were combined in a 1:1 ratio in H-buffer (750 mM NaCl, 125 mM Tris pH 9, 100 mM EDTA), and hybridized in a thermocycler (99°C for 3 min, and two cycles of 65°C for 5 min, and 30°C for 15 min). Hybridized products were transformed into chemically competent DH5α *E. coli* and plated on LB-ampicillin agar. Clones were sequenced and used for a second SLIM procedure to generate Fn69 or Fn61 FG loop regions. Bacterial expression was performed as described for Fn16.3 expression and purification.

Table 1. Oligos used in library construction and mRNA display

FnOligo1[a]	TTCTAATACGACTCACTATAGGGACAATTACTATTTACAATTACAATG CTCGAGGTCAAGG
FnOligo2[a]	CAATTACAATGCTCGAGGTCAAGGAAGCATCACCAACCAGCATCCAGATCAGCTGG
FnOligo3[ab]	ACCAGCATCCAGATCAGCTGGNNSNNSNNSNNSNNSVTTCGCTACTACCGCATCACCTACG
FnOligo4	GCACGGTGAATTCCTGGACAGGGCTATTGCCACCAGTTTCACCGTAGGTGATGCGGTAGTAGC
FnOligo5	CCTACCGGTCTCAGCTGATGGTAGCAGTGGACTTGCTGCCAG
FnOligo6	CCTACCGGTCTCACAGCGGCCTGAAACCTGGTGTCGACTATACCATCACGGTGTACGCCGTCACG
FnOligo7[b]	CGGTAGTTGATGGAGATCGGSNNSNNSNNSNNSNNSNNSNNSNNSNNCGTGACGGCGTACA CCGTGA
FnOligo9[c]	GGAGCCGCTACCCTTATCGTCGTCATCCTTGTAATCGGATCCGGTGCGGTAGTTGATGGAGATCG
FnOligo10[c]	GGAGCCGCTACCCTTATCGTCG
Splint1[c]	TTTTTTTTTTTTGGAGCCGCTACC
pF30P	5'-phospho-A_{21}-9_3-ACC-Pu (9= phosphoradmidite spacer 9, Pu= puromycin)
FnOligo9.2[d]	GCTTCCACTTCCGGACTTGTCATCGTCATCCTTGTAATCGGATCCGGTGCGGTAGTTGATGGAGAT CGG
FnOligo10.2[d]	GCTTCCACTTCCGGACTTGTCA
Splint2[d]	TTTTTTTTTTTTGCTTCCACTTCC

[a] Modified oligos that incorporate solubilizing mutations

[b] N denotes mixture of 20% T and G, 30% A and C. S denotes mixture of 40%G, 60% C

[c] denotes primers used during rhALCAM selection

[d] denotes primers used during vALCAM selection

Table 2. Cloning primers

ALCAM-For	CCACCAAGAAGGAGGAGGA
ALCAM-V1V2-Rev	AAATATCAAATACTGCCTGTTCAGA
ALCAM-BamHI-For	GAGAGAGGATCCACCAAGAAGGAGGAGGA
ALCAM-V1V2-EcoRI-Rev	GGAAGGAATTCGAAATATCAAATACTGCCTGTTCAGA
EcoRV-5138-F	GAGAGATATCACATATGCTCGAGGTCAAGGAAGCA
Fn-BamHI-5140-R	GGTGGTGGATCCGGTGCGGTAGTTGATG
Fn-BpmI-For	AGGAAGCTGGAGCAACCAGCATC
Fn-BpmI-Rev	TGATGGGAGGTCGTGCGGTAG
67-BC-Ft	AGCTTGCAGCCGTTGGTTCGCTACTATCGCATCACCTAC
67-BC-Fs	GTTCGCTACTATCGCATCACCTAC
67-BC-Rt	CAACGGCTGCAAGCTGGTCCAGCTGATCTGGATGCT
67-BC-Rs	GGTCCAGCTGATCTGGATGCT
67-FG-Ft	CCCAACTGGCCGTGGCACAACTACCCGATCTCCATCAACTACCG
67-FG-Fs	AACTACCCGATCTCCATCAACTACCG
67-FG-Rt	GTGCCACGGCCAGTTGGGCCAGAACGTGACGGCGTACACC
67-FG-Rs	CCAGAACGTGACGGCGTACACCGTG
61-BC-Ft	GAGCCCTACAGCTTCGTTCGCTACTATCGCATCACCTAC
61-BC-Fs	GTTCGCTACTATCGCATCACCTAC
61-BC-Rt	GAAGCTGTAGGGCTCGGGCCAGCTGATCTGGATGCTG
61-BC-Rs	GGGCCAGCTGATCTGGATGCTG
61-FG-Ft	GACTCGTCCAGCCCCTTCTTGCGGCCGATCTCCATCAACTACCG
61-FG-Fs	TTGCGGCCGATCTCCATCAACTACCG
61-FG-Rt	GAAGGGGCTGGACGAGTCGTACCACGTGACGGCGTACACCGTG
61-FG-Rs	GAAGGGGCTGGACGAGTCGTACCACGTGACGGCGTACACCGTG

RESULTS

e10FnIII library selection targeting full length ALCAM identifies a dominant clone Fn16.3

mRNA display was performed using two independently synthesized e10FnIII libraries to find ALCAM specific binding proteins. The first selection used a full length commercial human ALCAM, rhALCAM (R&D Systems) as the target ligand. The number of PCR cycles between selection cycles was monitored by agarose gel electrophoresis to prevent overamplification and skewing of the library. Library convergence was determined by a drop in the number of PCR cycles.

mRNA display was performed with continuous flow magnetic separation (CFMS) (6) for 4 rounds, after which affinity enrichment was performed in duplicate with either CFMS (Pool 4A) or with protein G-sepharose loaded beads (Pool 4B). Subclones from both pools were individually sequenced and screened for potential binders (Fig. 4).

Twenty clones sequenced from pool 4A were unique with 6 sequences judged as acceptable candidates based on analysis of primary structure. Out of 13 clones from pool 4B, one sequence that appeared in duplicate was also present in pool 4A. Pool 4A was subject to a fifth round of selection. All 15 sequences cloned from round 5 corresponded with a sequence found in both pools 4A and 4B. This clone is to be referred hereafter as Fn16.3.

Fn16.3 was found to have a lysine to asparagine (K57N) missense mutation at position 57 of the fibronectin backbone (marked by * in Fig. 4 under the reference sequence), but the polar residue substitution appeared to have no effect on binding efficiency (results not shown). Because Fn16.3 was identified in both round 4 pools, and dominated the round 5 pool, the single clone was considered the most promising binder to proceed with expression analysis in bacteria.

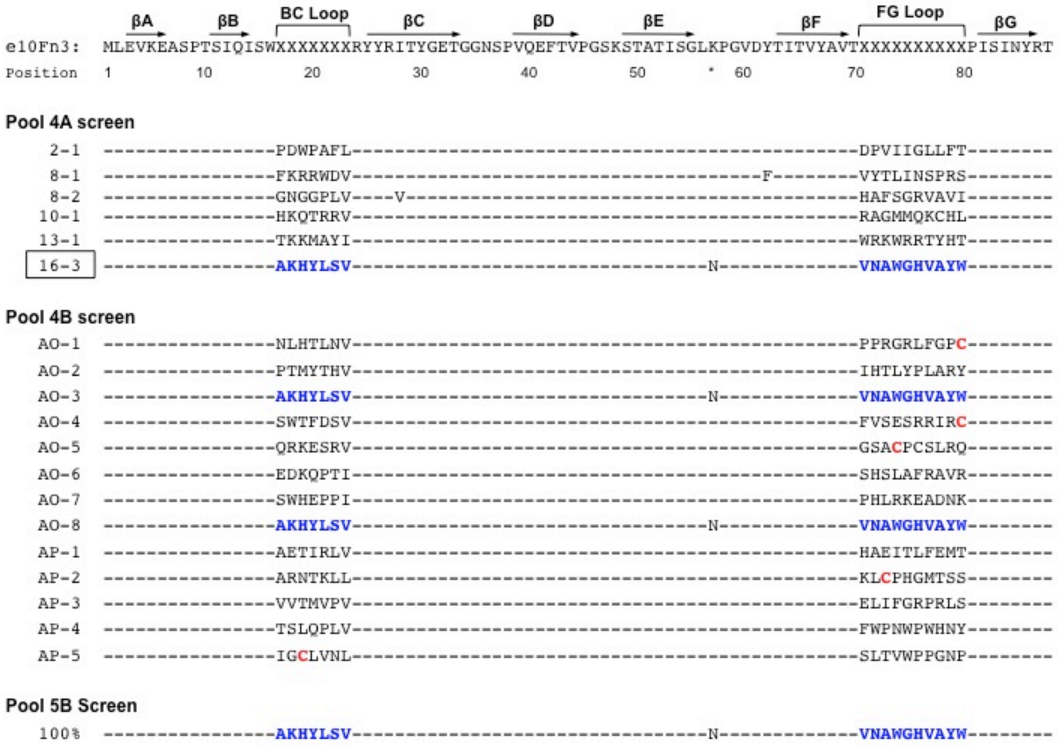

Figure 4. Loop sequences of e10FnIII variants after 4 rounds of selection. Out of 20 sequenced clones from pool 4A, six potential binders are shown, three of which have backbone substitutions. Pool 4B contains five clones with cysteine residues in randomized loops (highlighted red). All 15 clones screened after round 5 selection were identical to a clone (Fn16.3) found in both 4A and 4B pools (highlighted blue).

Bacterial expression of Fn16.3 does not yield a soluble product

mRNA-display utilizes an *in vitro* translation step with reticulocyte lysate which precludes knowing how a specific variant will behave when expressed in a bacterial host. Ideally, variants will be robustly expressed as soluble fractions but this is not always the case.

In order to evaluate bacterial expression, Fn16.3 cloned into a pET-11 derived vector, pAO5, which encodes a C-terminal 6xhis tag. pAO5-Fn16.3 was transformed into *E. coli* BL21(DE3) and induced under standard conditions (500 μM IPTG at 37°C for 3 hours).

Insolubility of Fn16.3 proved to be a persistent problem. Even after clearing lysates of insoluble material, continuing precipitate formation interfered with purification by severely impeding or completely blocking flow rates through Ni-NTA IMAC columns. SDS-PAGE analysis of purification fractions (Fig. 5) showed aggregates trapped in the stacking gel. While there appears to be a faint band present at Fn16.3's expected size (~11 kDa), the majority of protein is clearly expressed as insoluble particulates.

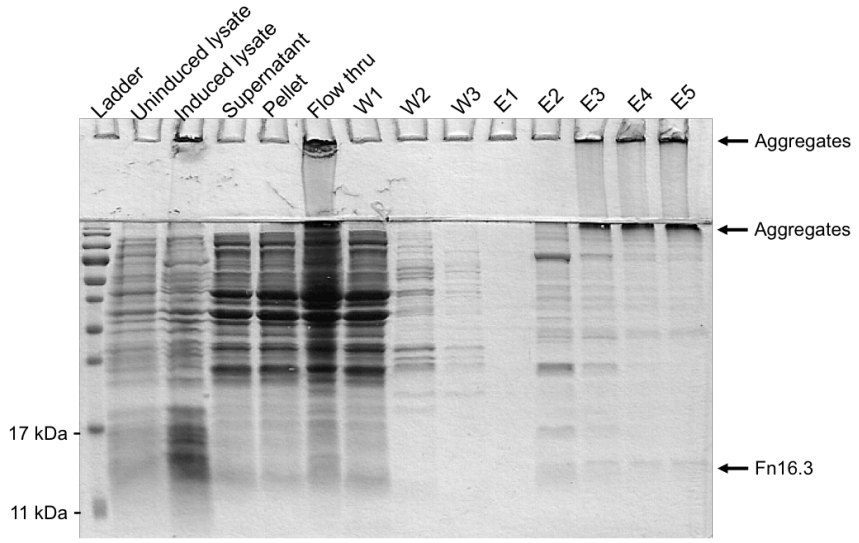

Figure 5. Fn16.3 purification via nickel-NTA affinity chromatography. Arrows indicate locations areas of trapped aggregates and expected molecular weight of Fn16.3 product.

Fn16.3 tagged with maltose binding protein (MBP) is partially soluble but non-functional

Out of the various strategies employed to optimize purification of prone-to-aggregate proteins, two well-described techniques were used to enhance recovery of soluble Fn16.3. The first was to fuse Fn16.3 to maltose binding protein (MBP), a solubility enhancing tag. The second was to adjust expression conditions to slow the rate of expression and protein folding by (a) lowering the IPTG inducer concentration and/or (b) lowering the induction temperature.

Fn16.3 was subcloned from pAO5 into pAO9, a modified pET28 vector that C-terminally appends an expression product with three tags: flag, maltose binding protein (MBP) and 6xhis. The large 44 kDa MBP is a powerful solubility enhancer that can also be exploited as an affinity tag for purification. pAO9-Fn16.3-MBP was expressed and IMAC purified in similar fashion to pAO5-Fn16.3, but incorporated an second purification step using amylose-agarose resin to capture MBP-tagged proteins. Purification fractions were analyzed via SDS-PAGE and Coomassie staining (Fig. 6).

Though there was still substantial aggregation in the total cell lysate, supernatant and nickel-NTA elution, a ~56 kDa product corresponding to the expected fusion product was found. The most concentrated amylose purified elution fraction contained high molecular weight aggregates and a ~110 kDa band that may correspond to a dimerized product. This purification indicated that a fraction of soluble protein is present, but pull down experiments indicated that MBP-solubilized fractions do not retain ALCAM binding activity (Fig. 7A-B).

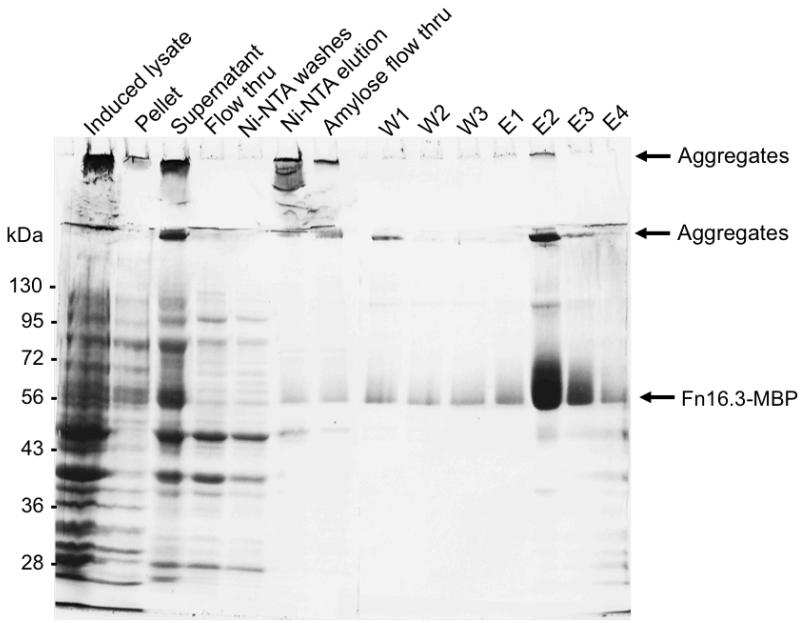

Figure 6. Fn16.3-MBP purification by two-step nickel-NTA and amylose-agarose affinity chromatography. Arrows indicate locations of aggregate formation and Fn16.3-MBP.

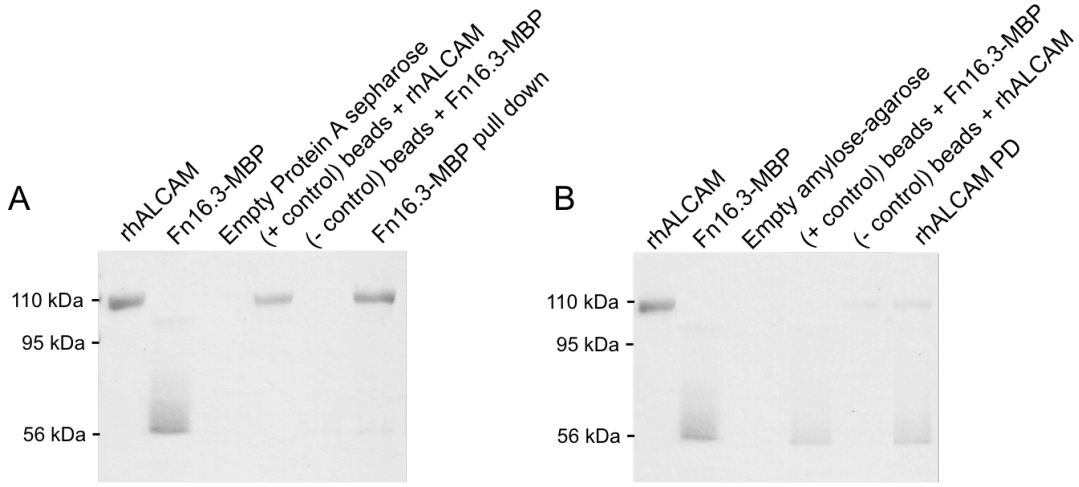

Figure 7. Pull down assays testing Fn16.3-MBP binding activity. (A) rhALCAM immobilized on protein A sepharose tested for Fn16.3-MBP binding in solution. (B) Fn16.3-MBP immobilized on amylose-agarose tested for rhALCAM pull down.

Lowered IPTG concentration and induction temperature does not improve Fn16.3 solubility

pET vectors (Merck/EMD) use the T7 polymerase promoter and lac operator to overexpress recombinant proteins at rates many times higher than bacterial RNA polymerase. With accelerated transcription levels, IPTG concentration and induction temperature can exacerbate misfolding of nascent proteins (28,29). Reducing IPTG concentrations and lowering induction temperature are common modifications to slow down expression and can enhance solubility for some proteins that are insoluble when expressed at standard conditions.

To evaluate whether these adjustments improve Fn16.3 solubility, Fn16.3 and Fn16.3-MBP cultures were expressed under varying induction conditions (Table 3). 1 ml cultures were lysed with SDS loading dye and assessed by SDS-PAGE and Coomassie staining. No significant change in the formation of aggregates was observed in Fn16.3 and Fn16.3-MBP expression cultures (results not shown).

Table 3. IPTG and temperature variations used in Fn16.3 and Fn16.3-MBP expression

Temperature	[IPTG]	Duration
37°C	1 mM	3 hrs
	0.5 mM	
	0.25 mM	
	0.1 mM	
	0.05 mM	
20°C	0.5 mM	3-4 hrs
		20-24 hrs
	0.05 mM	3 hrs
		Overnight (16 hrs)

Eukaryotic expression systems produce soluble Fn16.3 that binds ALCAM

In prokaryotic expression systems, transcription and translation occurs simultaneously, and certain amino acid sequences may be more prone to poor or improper folding by this process. Unsuccessful purification, likely caused by presence of insoluble aggregates, indicates that a

bacterial expression system does not accommodate proper folding and/or stability of Fn16.3. Because the Fn16.3 clone that binds ALCAM during selection is a product of *in vitro* translation in rabbit reticulocyte lysate, Fn16.3 was presumed to be soluble when expressed in eukaryotic hosts. However, to verify this assumption, Fn16.3 was expressed *in vitro* as well as in mammalian cells, and binding ability was tested with ALCAM pull down experiments.

Epitope-tagged Fn16.3 was expressed in with reticulocyte lysate and purified with M2-flag agarose. Recombinant ALCAM that is covalently attached to an Ig Fc domain (rhALCAM) was used as a target. rhALCAM immobilized on protein G sepharose was incubated with Fn16.3 and compared with protein G beads without rhALCAM. Anti-Flag immunoblot indicated Fn16.3 was rhALCAM specific, and not a result of non-specific binding to protein-G sepharose (Fig. 8)

In order to achieve expression in mammalian cell culture, Fn16.3 was cloned into the p3XFLAG-CMV-14 vector encoding a C-terminal flag-tag. The resultant p3XFLAG-Fn16.3 construct was transiently transfected into 293T cells. Fn16.3 was detected in whole cell lysate and purified via M2-Flag agarose (Fig. 9). vALCAM was incubated with Flag-tagged Fn16.3 bound to M2-Flag agarose or empty M2-agarose. Anti-ALCAM immunoblot indicated that vALCAM bound specifically to Fn16.3 (Fig. 10).

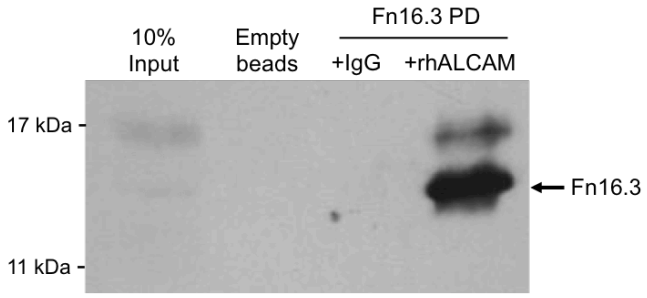

Figure 8. e10FnIII clone Fn16.3 selectively binds rhALCAM. Fn16.3 was incubated with either control protein G sepharose or rhALCAM loaded beads. Immunoblot was probed with anti-Flag.

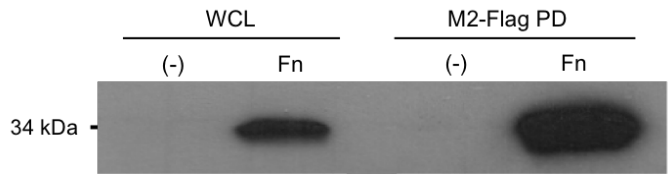

Figure 9. Fn16.3 is expressed in mammalian cells. 293T cells were either mock (-) transfected or transfected with p3XFLAG-Fn16.3. Recombinant Fn16.3 was purified with M2-Flag agarose. Immunoblot was probed with anti-Flag.

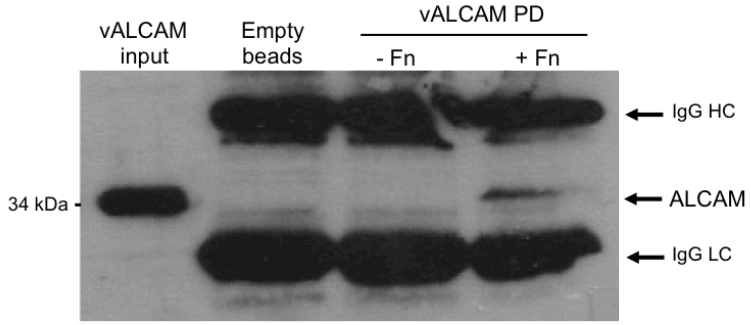

Figure 10. Fn16.3 expressed in mammalian cells binds vALCAM in solution. 250 ng input vALCAM was included in immunoblot for reference. vALCAM was incubated with either unloaded M2-Flag agarose or M2-Flag beads incubated with 293T lysates transfected with p3XFLAG-Fn16.3. Immunoblot was probed with anti-ALCAM.

The ideal e10FnIII variant binds specifically to a target ligand, but is also robustly produced in bacteria as a soluble protein. Although Fn16.3 met the first criteria, it was impractical to carry forward with more measured characterization of binding properties because soluble Fn16.3 would require production via mammalian cells.

In order to find a different ALCAM binder, we decided to construct a new e10FnIII library and repeat selection for two reasons. First of all, Fn16.3 was the only enriched clone identified by screening pools 4A, 4B and 5; all other clones were unique and unlikely to be functional binders.

Secondly, because Fn16.3 completely dominated the selection pool by round 5, we wanted to use a library where Fn16.3 did not already have a selection advantage.

The new mRNA display procedure differed in several respects. Importantly, new reverse transcription and PCR primers were designed to prevent contamination and amplification of rhALCAM selection sequences like Fn16.3. In addition, a truncated ALCAM consisting only of the V1-V2 domains was used as a selection target.

Cloning and expression of recombinant V1-V2 ALCAM domains (vALCAM)

The structural and functional characteristics of native ALCAM dictated how recombinant target protein was generated. From a functional perspective, ligand binding ability is lost in ALCAM deletion mutants lacking N-terminal V1-V2 domains. A binding protein that specifically interacts with one or both of these domains could be useful for disrupting ALCAM mediated cell-cell interactions. As a result, ALCAM was cloned to generate a truncated target for selection with 10FnIII.

From a structural perspective, vALCAM was generated as a secreted protein in a mammalian system because vALCAM's Ig-like domains require post-translational glycosylation and disulfide bridge formation. The V1-V2 domains (referred to as vALCAM) were cloned into pJD1, a modified pcDNA3 vector that encodes a 6xhis tag for affinity purification and a biotin ligase recognition site. pJD1-vALCAM was transfected into 293T cells and harvested media was purified by nickel-NTA affinity chromatography.

vALCAM was expressed at low levels in 293T cells, yielding an average of 10-15 μg from 30 ml of media harvested over 3 days. SDS-PAGE and immunoblot analysis verified expression and purification of the ~34 kDa soluble protein (Fig. 11A-B).

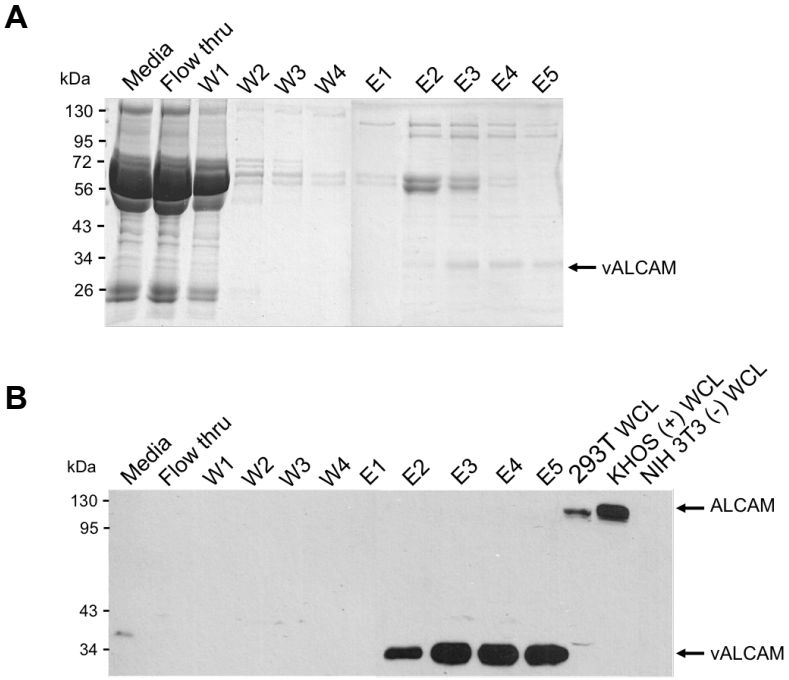

Figure 11. vALCAM purification analysis. (A) Coomassie stain and (B) Anti-ALCAM immunoblot. 100 μg cell lysate control samples were run alongside purification fractions.

e10FnIII library selection with vALCAM results a single binder identical to Fn16.3

After six rounds of selection 100% of 16 sequenced clones matched the Fn16.3 variant isolated from rhALCAM selection. Earlier pools were also screened for alternate clones, but 29/30 sequences sampled from pools 2-4 were also identical to Fn16.3. The single exception was a clone from pool 3 that differed from Fn16.3 only in the FG loop sequence.

Incorporating PmlI treatment in vALCAM selection fails to identify additional e10FnIII variants

Alternate e10FnIII variants specific to vALCAM could not be identified in sequencing screens due to overrepresentation of the Fn16.3 clone. The persistence of Fn16.3, even after precautionary measures to prevent sequence cross over from the first rhALCAM selection, required that we rethink our strategy of finding an alternate ALCAM binder. Fortuitously, a restriction site specific to the Fn16.3 FG loop was identified. Treating selection pools with PmlI, a type I enzyme that recognizes the 5'-CACGTG-3' sequence encoded in Fn16.3's FG loop, provided the opportunity to clear away Fn16.3 specific sequences and evaluate a more diverse pool that hopefully contained a different vALCAM binder.

By incorporating digestion prior to transcription, any partially transcribed Fn16.3 would lack the 3' sequence required for splint ligation and subsequently be removed during oligo(dT) purification. Assuming 100% cleavage, Fn16.3 mRNA-protein fusions should be absent during affinity enrichment, allowing for the amplification of weak or lower copy number binders.

Archived DNA (~10%) was subject to PmlI digestion to first assess Fn16.3 representation per round of selection (Fig. 12). The proportion of Fn16.3 steeply increased between rounds 2 and 3. Pool 3 was chosen for re-selection with PmlI, with the rationale that it would be more likely to contain variants in competition with Fn16.3 than clones in pool 2. Three additional rounds of selection were required to achieve pool convergence, after which six clones were sequenced and found to be identical to Fn16.3. Nucleotide sequence analysis revealed point mutations within the PmlI restriction site (Fig. 13), allowing these mutants to evade cleave and re-populate the selection pool during amplification.

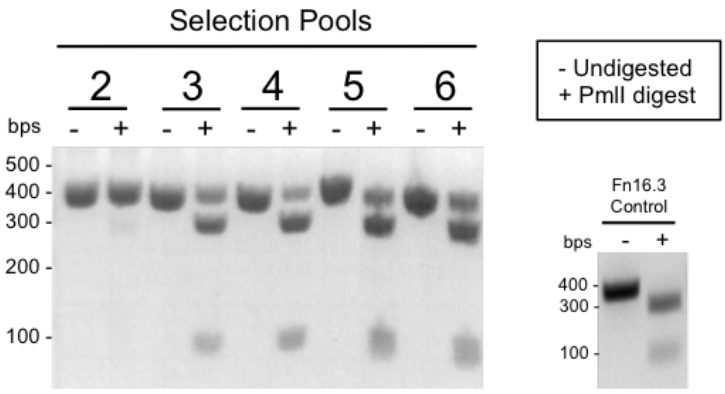

Figure 12. PmlI restriction digest of vALCAM selection pools 2-6. 1 µg of archived DNA was digested with PmlI. Control digest (bottom right panel) of single Fn16.3 species demonstrates efficient digestion.

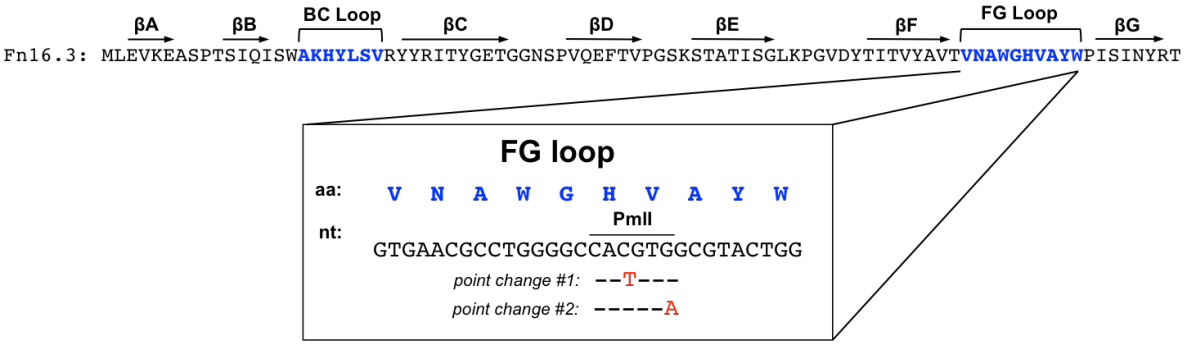

Figure 13. FG loop point mutations identified in pool 5P Fn16.3 clones. Sequence analysis revealed point mutations located within the PmlI recognition sequence.

High throughput sequencing (HTS) analysis of three selection pools

Clearly, repeating selection with a new library and treating selection pools with PmlI failed to identify non-Fn16.3 ALCAM binders. In fact, these two approaches failed to identify any non-Fn16.3 clones at all. However, the small-scale sequencing screens (usually 10-100+ sequences) used to identify enriched clones affords only a restricted view into complex library pools. Overrepresented clones like Fn16.3 make it difficult to identify alternate variants that can bind

ALCAM, albeit at lower affinities. High throughput sequencing (HTS), which can be used to rapidly analyze millions of sequences for potential variants, was used to evaluate the diversity of three e10FnIII selection pools that converged on Fn16.3. rhALCAM selection pool 4B was chosen to represent the original library where Fn16.3 was identified. vALCAM pools 3 and 5P (PmlI digested) were used to assess whether PmlI treatment had any effect on the enrichment of non-Fn16.3 binders.

Archived DNA from these libraries was amplified with customized primers for Illumina sequencing. Samples submitted for processing by the Illumina Genome Analyzer IIx were read by paired end sequencing. The three pools contributed to ~10% of a sequencing lane, which can theoretically return 20 million sequence reads. Raw data was analyzed by software that ranked amino acid sequences by copy number.

Pool 4B had the most number of unique sequences as well as a greater spread in copy numbers (Fig. 14). The two highest ranked sequences, with 69 and 61 copies, were previously unidentified in sequencing screens. Fn16.3 ranked first in vALCAM pools 3 and 5P. One clone (PEPYSFV-WYDSSSPFLR) appeared in all three pools and was the second ranking sequence in pool 4B. Reducing filtering stringency yielded no significant change in enrichment patterns except that the Fn16.3 clone surfaced as the top hit in pool 4B (data not shown).

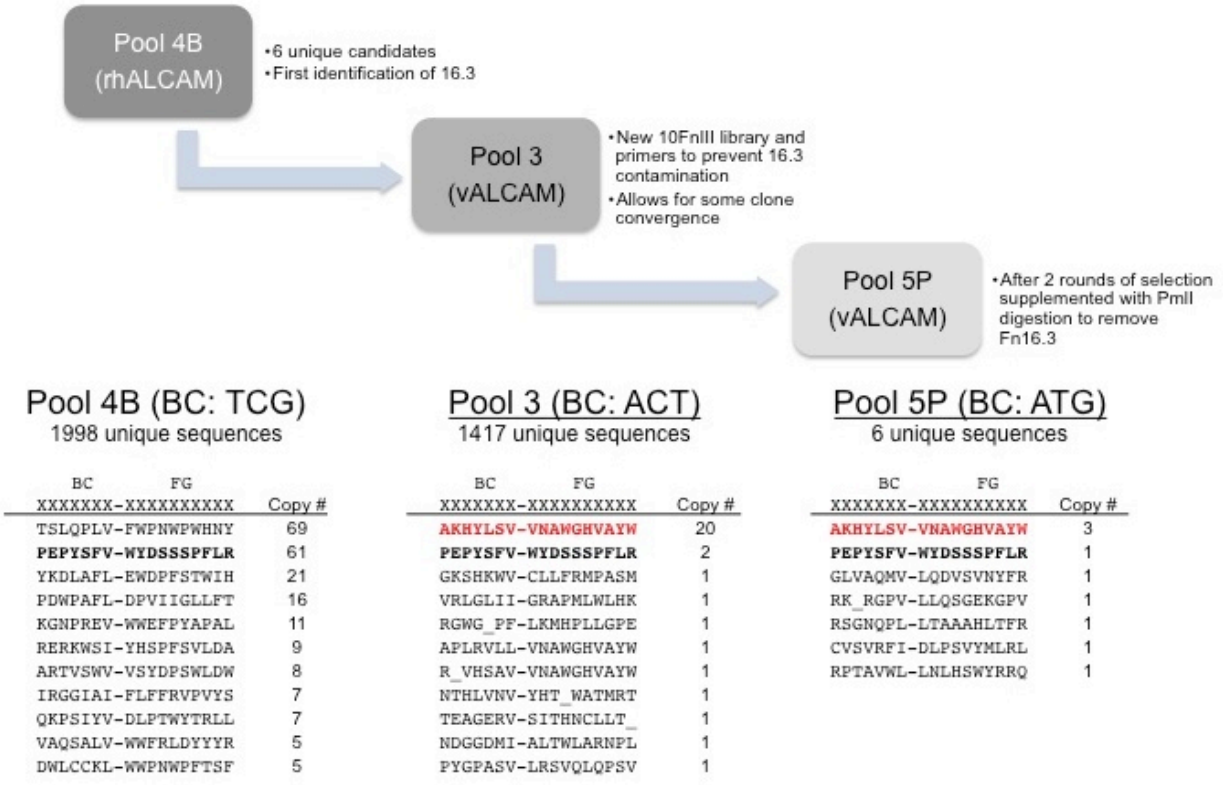

Figure 13. Selection pools sampled for HTS and ranked by clone enrichment. Data mining filtered sequences by pool specific barcodes (BC). Figure shows only BC and FG randomized regions with unique sequences ranked by copy number. 1998 unique sequences were identified in pool 4B, with the most enriched sequence appearing 69 times. 1417 unique sequences were identified in pool 3, and Fn16.3 was the most enriched with 20 copies. 6 unique sequences were identified in pool 5P, with Fn16.3 (red) appearing 3 times.

Enriched clones identified by HTS are expressed as soluble fractions in bacteria but lack ALCAM specific binding activity

HTS identified two new clones in rhALCAM pool 4B that are enriched relative to other sequences. One of these sequences (PEPYSFV-WYDSSSPFLR) appeared across all three pools. As a result, these two sequences were found to be the most promising binders based on preliminary HTS findings and were cloned to test for bacterial expression and ALCAM binding.

Because these variants were identified via HTS, Fn61 and Fn69 sequences (designated by copy number in pool 4B) could not be simply cloned into an expression vector. The two clones were generated synthetically by SLIM-PCR, an inverse PCR method that used oligos specific to Fn69 and Fn61 to replace Fn16.3's BC and FG regions in the pAO5-Fn16.3 plasmid. Both clones were expressed in *E. coli* BL21(DE3) as soluble fractions (Fig15, A-B), but neither exhibited ALCAM specific binding activity (Fig 16, A-B).

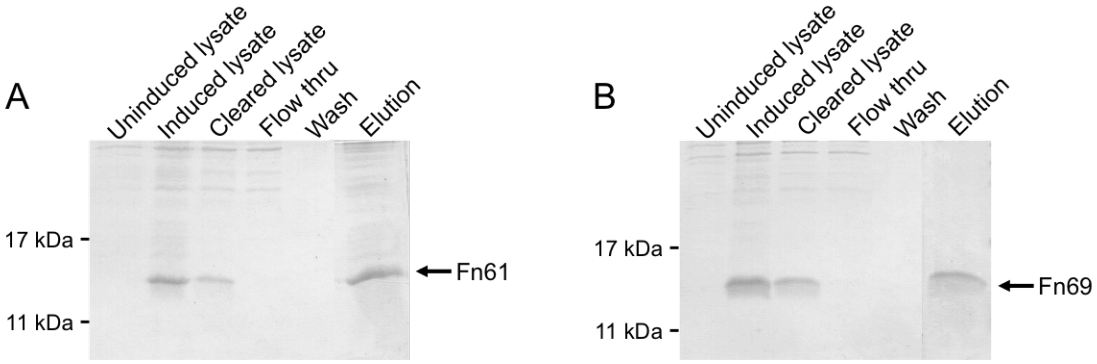

Figure 14. Nickel-NTA purifications of (A) Fn61 and (B) Fn69. Coomassie stained gels show distinct bands corresponding to Fn variants are seen in lysates and elution fractions only.

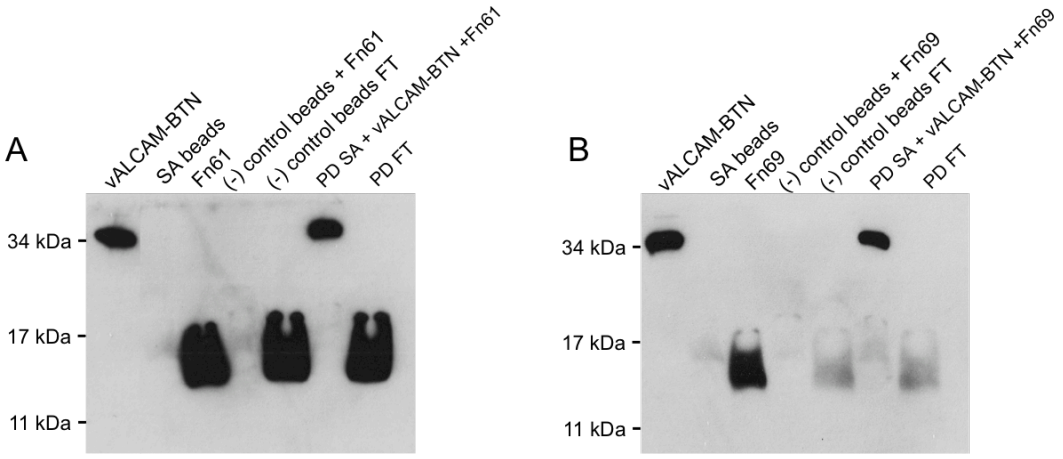

Figure 15. Pull down assay to test ALCAM binding of (A) Fn61 and (B) Fn69. vALCAM was conjugated to streptavidin agarose and combined in solution with Fn clones. Immunoblots were probed with anti-His.

DISCUSSION

mRNA display results in convergence on a single ALCAM binder, Fn16.3

mRNA display was successfully used to identify an e10FnIII clone that binds to ALCAM, named Fn16.3. Unfortunately, the inability to recover sufficient qualities of soluble Fn16.3 from a bacterial host derailed initial prospects of conducting additional binding studies.

Our mRNA display selection targeted two isoforms of human ALCAM; full length rhALCAM that consists of all 5 extracellular domains and truncated vALCAM which is composed by the 2 membrane distal V1-V2 domains. Both selections converged on a single dominant binder, Fn16.3. This indicates that Fn16.3 recognizes an epitope on within the V1-V2 region, which is also involved in antigen recognition at the cellular level.

Expression and purification

Ideally, recombinant proteins will be soluble, stable, and expressed at high levels. Realistically, most expression conditions or the recombinant molecules themselves need to be optimized and/or modified on a case-by-case basis.

When recombinant proteins are poorly expressed or expressed in inclusion bodies, there are various tactics available to try and coax an unruly protein into soluble fractions. In the case of Fn16.3, two traditional modifications were made; fusion to a solubility enhancing tag and modification to induction conditions. Only these two approaches were used because a goal of this project was to isolate a robustly expressed protein that can be purified in an uncomplicated fashion.

MBP is a popular fusion tag because it dually acts as an affinity tag and a solubility enhancer, but can enhance solubility to the extent that misfolded target proteins are pulled into soluble fractions (28). This may explain why it was possible to purify a soluble fraction of Fn16.3 fusion product that did appear to bind ALCAM. Typically, 15-18°C is the lower limit when reducing induction temperature in *E. coli* strains without cold adapted chaperones. Recently, recombinant mannanase and cellulase proteins that form inclusion bodies at 15-37°C have been solubly expressed when cultured at 6-10°C in unmodified BL21(DE3) cells (30). Thus, a more drastic reduction in growth temperature may further slow down protein synthesis. Other studies have found that a multitude of additives may improve solubility for a range of proteins. One popular additive is L-arginine, which has been shown to suppress protein aggregation (31). Even with modifications to expression protocols, very pure protein samples can precipitate even at low concentrations (less than 1 mg/ml) (32).

Extensive testing of expression conditions was decided to be a time consuming and impractical process (33) that does not guarantee resolution of expression difficulties. More importantly, one facet of the project was to identify a stable protein with an uncomplicated purification procedure, not to be encumbered by a fickle protein. Previously isolated 10FnIII and e10FnIII have been shown to be stable and expressed at levels ranging from 4-20 mg/L (12). As a result, the issue of solubility is likely an inherent property to Fn16.3's loop regions.

Expression in eukaryotic vs. prokaryotic systems

Eukaryotic cells compartmentalize transcription and translation and include chaperone proteins that facilitate protein folding. In contrast, transcription and translation in prokaryotes

occurs simultaneously at much faster rates. In addition, there is the issue of codon bias, where rare codons can deleteriously affect heterologous protein expression. Positioning of high and low frequency codons can be influential in pacing the rate of translation and proper folding of nascent proteins (34). Codons encoding the e10FnIII backbone were selected for optimized expression in mammalian and bacterial cultures, but is unaccounted for in the randomized loop regions.

Contributing factors to the limited diversity in e10FnIII library used for vALCAM selection

One matter of concern from the vALCAM selection in particular was the apparent homogeneity of the e10FnIII library used for vALCAM selection. It is clear that the Fn16.3 clone quenched the diversity of the selection pool, but PmlI digestion indicated that the clone represented a large fraction of sequences following two rounds of selection. It is unlikely that Fn16.3 clones from rhALCAM-targeted selection contaminated the new library because a separate primer set was used during amplification reactions. However, there are two discrepancies from this particular selection that may have contributed to early curtailing of library complexity.

First, utilizing negative selection in round 1 of selection may have inadvertently eliminated weak binders. Because synthesis of the high complexity library yields a large number of unique variants represented by a low copy number (<10 copies) (12) stringent binding conditions can rapidly reduce library complexity in early rounds. As a result, negative selection and Flag preselection are often implemented in round 2 or later.

Secondly, inefficient transcription during round 2 of selection may have winnowed out potentially important variants. There was a steep drop in transcription yield in round 3, <25% than that of subsequent rounds (Table 4). Poor transcription early in selection could have significantly reduced the concentration of transcripts capable of forming mRNA-linker fusions products.

In a broader sense, it is important to consider that the degree of clone enrichment is not determined by binding affinity alone. The *in vitro* selection process depends on four enzymatic steps (DNA amplification, transcription, ligation, translation), and one or more of these processes can influence how well a clone is represented in a sequencing screen. In particular, the exponential rate of at which PCR amplification occurs can rapidly imbalance a library pool. Preferential amplification caused by small clonal differences like tandem repeats or melting temperature can over or underrepresent clones that otherwise have similar binding properties (35). Likewise, it *in vitro* transcription and translation reactions may have their own biases for specific sequences or codons. Even if a clone can be efficiently transcribed and translated, improper fusion between mRNA and protein molecules renders them useless. Ultimately, it is a combination of these factors that determines the final pool of mRNA-protein fusions that can be selected for functional activity. Fn16.3 clearly emerged as the rhALCAM and vALCAM selection "winner," but for the reasons explained above we cannot be discount the notion that other, potentially stronger and more soluble binders, were also present in the e10FnIII libraries.

Table 4. Transcription yields and PCR cycles for library selection with vALCAM

Round #	Transcription yield (pmoles)	# PCR cycles
1	928	17
2	334	21
3	1496	26
4	1450	22
5	1277	20
6	1452	16

Further evolution of Fn16.3 and selecting for solubilizing mutations

Although Fn16.3 cannot be appreciably expressed in bacteria, Fn16.3 expressed in eukaryotic cells does bind specifically to ALCAM. As a result, Fn16.3 can be further evolved i.e., introducing mutations by error prone PCR, and re-selecting variants for increased solubility. Error prone PCR is useful in that salt concentrations can be adjusted to control the rate at which mutations are introduced into a sequence.

There are several methods that can select mutants with increased solubility properties. Béhar et al. combined error prone PCR with selection by fusion to *chloramphenicol acetyltransferase (CAT),* to evolve interleukin-15 (IL-15); IL-15, which normally forms inclusion bodies when expressed in bacteria, was evolved into a soluble and functional variant (36). CAT-fusions are useful because solubility of the fused protein influences CAT's ability to confer resistance to chloramphenicol (Cam); by plating variants on high Cam media, clones with increased solubility can be isolated and further characterized for binding activity (37). Van den Verg et al. used error prone PCR and fusions to green fluorescent protein (GFP) to screen for enhanced solubility of the tobacco etch virus (TEV) protease (38). Olson et al. also used the GFP reporter method to improve expression of a phospho-IκBα specific e10FnIII variant in *E. coli* BL21(DE3) (14). One caveat of using the GFP reporter system, however, is that because fluorescence represents total soluble protein rather than the protein of interest alone, results may not accurately reflect an acquired function (12).

Another alternative to using error-prone PCR and reselection is combining mRNA and yeast displays to take advantage of the benefits endowed by each platform. By starting selection with mRNA display, a high complexity library can be screened for binding ability. After a

couple rounds of selections, many of the undesirable variants would be eliminated and the complexity of the pool reduced. With a significantly reduced complexity, the pool can be cloned and transitioned into the yeast-display platform, which can accommodate a complexity of ~10^8. This combined approach would allow for expressing variants in an *in vivo* system, better reflecting performance in physiological conditions. Moreover, yeast intrinsically tend to more robustly express variants with stable biophysical properties, and as such would enrich for clones that exhibit both binding ability and structural stability (39).

Additional remarks

mRNA display is an appealing technique for developing novel binding proteins because in the spectrum of display platforms, it marks the upper limit of library diversity currently available. However, this level of diversity is a quality needed for highly specific binding of obscure or unstructured targets. A lower complexity library may be sufficient for selection with a large signaling molecule like ALCAM that consists of multiple Ig V-type and C-type domains.

In recent years, the advancement of technology, particularly in the computing power of large datasets, has allowed the field of protein engineering to develop algorithms to predict protein behavior. Klus et al. developed the cleverSuite program analyze the physico-chemical properties of protein sequences to predict different features such as structural disorder, chaperone interactions and RNA-binding ability (40). The ccSOL *omics* webserver utilizes an algorithm that predicts solubility of both endogenous and heterologous proteins expressed in *E. coli* (41). Analytical programs such as these may provide supplemental insight when screening sequences for acceptable clones.

Conclusion

In conclusion, two separate selections on independently synthesized e10FnIII libraries led to the identification of a single clone, termed Fn16.3. When expressed *in vitro* and in a mammalian expression system, Fn16.3 selectively binds to full length and truncated versions of ALCAM, verifying that it binds specifically to the desired target. Fn16.3 expression in bacteria yields an insoluble product that cannot be rescued by an affinity enhancing tag like MBP. However, because Fn16.3 demonstrated specificity to ALCAM, it is a strong candidate for further evolution and selection for solubility enhancing mutations.

REFERENCES

1. Chames, P., Van Regenmortel, M., Weiss, E., and Baty, D. (2009) Therapeutic antibodies: successes, limitations and hopes for the future. *British journal of pharmacology* **157**, 220-233
2. Carter, P. J. (2006) Potent antibody therapeutics by design. *Nature reviews. Immunology* **6**, 343-357
3. Gebauer, M., and Skerra, A. (2009) Engineered protein scaffolds as next-generation antibody therapeutics. *Current opinion in chemical biology* **13**, 245-255
4. Smith, G. P., and Petrenko, V. A. (1997) Phage Display. *Chemical reviews* **97**, 391-410
5. Roberts, R. W. (1999) Totally in vitro protein selection using mRNA-protein fusions and ribosome display. *Current opinion in chemical biology* **3**, 268-273
6. Olson, C. A., Adams, J. D., Takahashi, T. T., Qi, H., Howell, S. M., Wu, T. T., Roberts, R. W., Sun, R., and Soh, H. T. (2011) Rapid mRNA-display selection of an IL-6 inhibitor using continuous-flow magnetic separation. *Angewandte Chemie* **50**, 8295-8298
7. Schiff, D., Kesari, S., de Groot, J., Mikkelsen, T., Drappatz, J., Coyle, T., Fichtel, L., Silver, B., Walters, I., and Reardon, D. (2014) Phase 2 study of CT-322, a targeted biologic inhibitor of VEGFR-2 based on a domain of human fibronectin, in recurrent glioblastoma. *Investigational new drugs*
8. Tolcher, A. W., Sweeney, C. J., Papadopoulos, K., Patnaik, A., Chiorean, E. G., Mita, A. C., Sankhala, K., Furfine, E., Gokemeijer, J., Iacono, L., Eaton, C., Silver, B. A., and Mita, M. (2011) Phase I and pharmacokinetic study of CT-322 (BMS-844203), a targeted Adnectin inhibitor of VEGFR-2 based on a domain of human fibronectin. *Clinical cancer*

research : an official journal of the American Association for Cancer Research **17**, 363-371

9. Emanuel, S. L., Engle, L. J., Chao, G., Zhu, R. R., Cao, C., Lin, Z., Yamniuk, A. P., Hosbach, J., Brown, J., Fitzpatrick, E., Gokemeijer, J., Morin, P., Morse, B. A., Carvajal, I. M., Fabrizio, D., Wright, M. C., Das Gupta, R., Gosselin, M., Cataldo, D., Ryseck, R. P., Doyle, M. L., Wong, T. W., Camphausen, R. T., Cload, S. T., Marsh, H. N., Gottardis, M. M., and Furfine, E. S. (2011) A fibronectin scaffold approach to bispecific inhibitors of epidermal growth factor receptor and insulin-like growth factor-I receptor. *mAbs* **3**, 38-48
10. Koide, A., Bailey, C. W., Huang, X., and Koide, S. (1998) The fibronectin type III domain as a scaffold for novel binding proteins. *Journal of molecular biology* **284**, 1141-1151
11. 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., and Lipovsek, D. (2002) Directed evolution of high-affinity antibody mimics using mRNA display. *Chemistry & biology* **9**, 933-942
12. Olson, C. A., and Roberts, R. W. (2007) Design, expression, and stability of a diverse protein library based on the human fibronectin type III domain. *Protein science : a publication of the Protein Society* **16**, 476-484
13. Liao, H. I., Olson, C. A., Hwang, S., Deng, H., Wong, E., Baric, R. S., Roberts, R. W., and Sun, R. (2009) mRNA display design of fibronectin-based intrabodies that detect and inhibit severe acute respiratory syndrome coronavirus nucleocapsid protein. *The Journal of biological chemistry* **284**, 17512-17520

14. Olson, C. A., Liao, H. I., Sun, R., and Roberts, R. W. (2008) mRNA display selection of a high-affinity, modification-specific phospho-IkappaBalpha-binding fibronectin. *ACS chemical biology* **3**, 480-485

15. Bowen, M. A., Patel, D. D., Li, X., Modrell, B., Malacko, A. R., Wang, W. C., Marquardt, H., Neubauer, M., Pesando, J. M., Francke, U., and et al. (1995) Cloning, mapping, and characterization of activated leukocyte-cell adhesion molecule (ALCAM), a CD6 ligand. *The Journal of experimental medicine* **181**, 2213-2220

16. Swart, G. W. (2002) Activated leukocyte cell adhesion molecule (CD166/ALCAM): developmental and mechanistic aspects of cell clustering and cell migration. *European journal of cell biology* **81**, 313-321

17. van Kempen, L. C., Nelissen, J. M., Degen, W. G., Torensma, R., Weidle, U. H., Bloemers, H. P., Figdor, C. G., and Swart, G. W. (2001) Molecular basis for the homophilic activated leukocyte cell adhesion molecule (ALCAM)-ALCAM interaction. *The Journal of biological chemistry* **276**, 25783-25790

18. Bowen, M. A., Aruffo, A. A., and Bajorath, J. (2000) Cell surface receptors and their ligands: in vitro analysis of CD6-CD166 interactions. *Proteins* **40**, 420-428

19. Ikeda, K., and Quertermous, T. (2004) Molecular isolation and characterization of a soluble isoform of activated leukocyte cell adhesion molecule that modulates endothelial cell function. *The Journal of biological chemistry* **279**, 55315-55323

20. Kijima, N., Hosen, N., Kagawa, N., Hashimoto, N., Nakano, A., Fujimoto, Y., Kinoshita, M., Sugiyama, H., and Yoshimine, T. (2012) CD166/activated leukocyte cell adhesion molecule is expressed on glioblastoma progenitor cells and involved in the regulation of tumor cell invasion. *Neuro-oncology* **14**, 1254-1264

21. Kahlert, C., Weber, H., Mogler, C., Bergmann, F., Schirmacher, P., Kenngott, H. G., Matterne, U., Mollberg, N., Rahbari, N. N., Hinz, U., Koch, M., Aigner, M., and Weitz, J. (2009) Increased expression of ALCAM/CD166 in pancreatic cancer is an independent prognostic marker for poor survival and early tumour relapse. *British journal of cancer* **101**, 457-464

22. Weichert, W., Knosel, T., Bellach, J., Dietel, M., and Kristiansen, G. (2004) ALCAM/CD166 is overexpressed in colorectal carcinoma and correlates with shortened patient survival. *Journal of clinical pathology* **57**, 1160-1164

23. Yan, M., Yang, X., Wang, L., Clark, D., Zuo, H., Ye, D., Chen, W., and Zhang, P. (2013) Plasma membrane proteomics of tumor spheres identify CD166 as a novel marker for cancer stem-like cells in head and neck squamous cell carcinoma. *Molecular & cellular proteomics : MCP* **12**, 3271-3284

24. McCabe, K. E., Liu, B., Marks, J. D., Tomlinson, J. S., Wu, H., and Wu, A. M. (2012) An Engineered Cysteine-Modified Diabody for Imaging Activated Leukocyte Cell Adhesion Molecule (ALCAM)-Positive Tumors. *Mol Imaging Biol* **14**, 336-347

25. Federman, N., Chan, J., Nagy, J. O., Landaw, E. M., McCabe, K., Wu, A. M., Triche, T., Kang, H., Liu, B., Marks, J. D., and Denny, C. T. (2012) Enhanced growth inhibition of osteosarcoma by cytotoxic polymerized liposomal nanoparticles targeting the alcam cell surface receptor. *Sarcoma* **2012**, 126906

26. Chiu, J., March, P. E., Lee, R., and Tillett, D. (2004) Site-directed, Ligase-Independent Mutagenesis (SLIM): a single-tube methodology approaching 100% efficiency in 4 h. *Nucleic acids research* **32**, e174

27. Chiu, J., Tillett, D., Dawes, I. W., and March, P. E. (2008) Site-directed, Ligase-Independent Mutagenesis (SLIM) for highly efficient mutagenesis of plasmids greater than 8kb. *Journal of microbiological methods* **73**, 195-198

28. Francis, D. M., and Page, R. (2010) Strategies to optimize protein expression in E. coli. *Current protocols in protein science / editorial board, John E. Coligan ... [et al.]* **Chapter 5**, Unit 5 24 21-29

29. Turner, P., Holst, O., and Karlsson, E. N. (2005) Optimized expression of soluble cyclomaltodextrinase of thermophilic origin in Escherichia coli by using a soluble fusion-tag and by tuning of inducer concentration. *Protein Expr Purif* **39**, 54-60

30. Song, J. M., An, Y. J., Kang, M. H., Lee, Y. H., and Cha, S. S. (2012) Cultivation at 6-10 degrees C is an effective strategy to overcome the insolubility of recombinant proteins in Escherichia coli. *Protein Expr Purif* **82**, 297-301

31. Ishibashi, M., Tsumoto, K., Ejima, D., Arakawa, T., and Tokunaga, M. (2005) Characterization of arginine as a solvent additive: a halophilic enzyme as a model protein. *Protein and peptide letters* **12**, 649-653

32. Structural Genomics, C., China Structural Genomics, C., Northeast Structural Genomics, C., Graslund, S., Nordlund, P., Weigelt, J., Hallberg, B. M., Bray, J., Gileadi, O., Knapp, S., Oppermann, U., Arrowsmith, C., Hui, R., Ming, J., dhe-Paganon, S., Park, H. W., Savchenko, A., Yee, A., Edwards, A., Vincentelli, R., Cambillau, C., Kim, R., Kim, S. H., Rao, Z., Shi, Y., Terwilliger, T. C., Kim, C. Y., Hung, L. W., Waldo, G. S., Peleg, Y., Albeck, S., Unger, T., Dym, O., Prilusky, J., Sussman, J. L., Stevens, R. C., Lesley, S. A., Wilson, I. A., Joachimiak, A., Collart, F., Dementieva, I., Donnelly, M. I., Eschenfeldt, W. H., Kim, Y., Stols, L., Wu, R., Zhou, M., Burley, S. K., Emtage, J. S., Sauder, J. M.,

Thompson, D., Bain, K., Luz, J., Gheyi, T., Zhang, F., Atwell, S., Almo, S. C., Bonanno, J. B., Fiser, A., Swaminathan, S., Studier, F. W., Chance, M. R., Sali, A., Acton, T. B., Xiao, R., Zhao, L., Ma, L. C., Hunt, J. F., Tong, L., Cunningham, K., Inouye, M., Anderson, S., Janjua, H., Shastry, R., Ho, C. K., Wang, D., Wang, H., Jiang, M., Montelione, G. T., Stuart, D. I., Owens, R. J., Daenke, S., Schutz, A., Heinemann, U., Yokoyama, S., Bussow, K., and Gunsalus, K. C. (2008) Protein production and purification. *Nature methods* **5**, 135-146

33. Lebendiker, M., and Danieli, T. (2014) Production of prone-to-aggregate proteins. *FEBS letters* **588**, 236-246

34. Angov, E., Legler, P. M., and Mease, R. M. (2011) Adjustment of codon usage frequencies by codon harmonization improves protein expression and folding. *Methods in molecular biology* **705**, 1-13

35. Walsh, P. S., Erlich, H. A., and Higuchi, R. (1992) Preferential PCR amplification of alleles: mechanisms and solutions. *PCR methods and applications* **1**, 241-250

36. Béhar, G., Solé, V., Defontaine, A., Maillasson, M., Quéméner, A., Jacques, Y., and Tellier, C. (2011) Evolution of interleukin-15 for higher E. coli expression and solubility. *Protein Engineering Design and Selection* **24**, 283-290

37. Maxwell, K. L., Mittermaier, A. K., Forman-Kay, J. D., and Davidson, A. R. (1999) A simple in vivo assay for increased protein solubility. *Protein science : a publication of the Protein Society* **8**, 1908-1911

38. van den Berg, S., Löfdahl, P.-Å., Härd, T., and Berglund, H. (2006) Improved solubility of TEV protease by directed evolution. *Journal of Biotechnology* **121**, 291-298

39. Koide, S., Koide, A., and Lipovsek, D. (2012) Target-binding proteins based on the 10th human fibronectin type III domain ((1)(0)Fn3). *Methods in enzymology* **503**, 135-156

40. Klus, P., Bolognesi, B., Agostini, F., Marchese, D., Zanzoni, A., and Tartaglia, G. G. (2014) The cleverSuite approach for protein characterization: predictions of structural properties, solubility, chaperone requirements and RNA-binding abilities. *Bioinformatics* **30**, 1601-1608

41. Agostini, F., Cirillo, D., Livi, C. M., Delli Ponti, R., and Tartaglia, G. G. (2014) ccSOL omics: a webserver for solubility prediction of endogenous and heterologous expression in Escherichia coli. *Bioinformatics* **30**, 2975-2977