DIRECT INTRACELLULAR SELECTION OF SYNTHETIC BINDING PROTEINS THAT SPECIFICALLY RECOGNIZE POST-TRANSLATIONALLY MODIFIED PROTEINS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Bunyarit Meksiriporn

December 2018

DIRECT INTRACELLULAR SELECTION OF SYNTHETIC BINDING PROTEINS THAT SPECIFICALLY RECOGNIZE POST-TRANSLATIONALLY MODIFIED PROTEINS

Bunyarit Meksiriporn, Ph. D.

Cornell University 2018

Post-translational modifications (PTMs), such as phosphate groups, control many cellular activities and play important roles in both health and disease. Accordingly, there is an urgent need for affinity reagents that target modified sites on individual proteins and can be used in applications ranging from immunodetection to immunotherapy. However, the generation of synthetic PTM-binders remains a major challenge using traditional hybridoma or selection methods, and the resulting affinity reagents are often nonspecific. Here, we describe a genetic selection strategy for routine laboratory isolation of phospho-specific designed ankyrin repeat proteins (DARPins) by linking *in vivo* affinity capture of a phosphorylated target protein with antibiotic resistance of *Escherichia coli* cells. The genetic assay was validated using an existing panel of DARPins that selectively bind the nonphosphorylated (inactive) form of extracellular signal-regulated kinase 2 (ERK2) or its doubly phosphorylated (active) form (pERK2). Using the selection strategy, we performed affinity maturation of a phospho-specific DARPin and uncovered several superior binders including a protein with 70-fold improved affinity (to 0.15 nM) for its cognate antigen, pERK2, but with no significant change in affinity for non-cognate ERK2. The selection strategy was similarly applied to reprogram the specificity of the same DARPin, yielding promiscuous variants that evolved the ability to bind non-cognate ERK2 but that still retained binding to pERK2. Collectively, these results establish our PTM-specific genetic selection as a useful and potentially generalizable new tool for studying PTM-specific binding proteins and customizing their affinity and selectivity.

BIOGRAPHICAL SKETCH

Bunyarit Meksiriporn was born and grew up in Bangkok, Thailand. His parents have always been supportive and have given him the best education they can provide. As the only child, Bunyarit has only his parents to look up to. Seeing his parents reading inspired his passion for reading at a young age, which has become an excellent foundation for his achievements today. He graduated first in his class in high school from Tepleela School, Bangkok, Thailand. Later, He was admitted to department of biochemistry, Faculty of Science, Chulalongkorn University which is the most prestigious university in Thailand. After he received her Bachelor of Science from Chulalongkorn University, Bangkok, Thailand. He was awarded National Innovation Agency (NIA) fellowship to pursue his study in microbiology major at Chulalongkorn University. After that, he was awarded Thai government fellowship to pursue Ph. D. in Biomedical Engineering at Cornell University in 2011. He joined DeLisa research group, and did his research on development of a novel protein selection technology against post-translationally modified proteins under supervision of his advisor, Dr. Matthew P. DeLisa.

To my father, Bunyong, for his love and support.

To my mother, Anong, for her endless love and care.

To all my friends at Cornell, for always being supportive for my tough Ph.D. life.

ACKNOWLEDGMENTS

First of all, I would like to take this opportunity to express my gratitude to my advisor, Dr. Matthew P. DeLisa, for his guidance, encouragement, support, and most importantly his patience. His extensive knowledge and optimism played an important role throughout my research. There were tough times when knowledge was not encouraging enough to keep me motivated, his encouragement was important in helping me achieve my PhD degree. This research project certainly would not have been possible without his advice and support. My gratitude is also due to my committee members, Dr. Jeffrey Varner and Dr. Hening Lin. I would like to thank all DeLisa Research Group (DLRG) members for being not only my labmates but also my friends, especially Didi (my first and foremost mentor who introduced me to the world of protein engineering), Alyse (my second mentor), Dr. Lee (my third mentor), Thapakorn (my lovely brother), May, Sai, Amy, Sean, Dom, and Laura. Their discussions are always valuable to me to troubleshoot my research problem. They were there through my ups and downs; laughing at my jokes and listening to my complaints. While working in my lab, I learned not only science but also English. I would like to thank all my students who involved in this project, including Allen, Kelvin, Raashed, and Mindy. I would like to acknowledge the funding provided by Royal Thai Government Scholarship throughout my Ph.D. I would also like to thank all my friends in Ithaca, including Chairat, Thapakorn, Narit, Sunsiree, Kittikun, Ithipong, Chinawat, Chalermpat, Siraphat, Weerin, Monchanok, Visarute, Ravi, Rawinthira, Natchaphon, and Chris for making Ithaca my home away from home. I would like to thank Thanawat and Teera who are my friends in Thailand and have been kindly giving me helpful advice, moral support, and sending me their love and care from afar. I am deeply thankful to Plengchart for his enduring love and caring. Ultimately, I would like to express my love and gratitude to my beloved family whose endless love and encouragement have always been a great part of my life.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

PROTEIN SELECTION TECHNOLOGY FOR POST-TRANSLATIONAL MODIFIED TARGETS

1. Introduction

In the last few decades, researchers have discovered that the number of human proteins is much more complex than that of human genes. It was estimated that the human genome theoretically encodes only 20,000– 25,000 protein-coding genes; however, the total number of proteins in the human proteome was estimated to be higher than 1 million (Fig. 1.1) [1]. These estimations of human proteins explicate that a single gene is able to encode multiple proteins. From the level of the genome to the transcriptome, cells utilize genomic recombination, transcription initiation at alternative promoters, differential transcription termination, and alternative splicing of the transcript as mechanisms rendering various mRNA transcripts from a single gene [2]. The increase in complexity and diversity from the level of the transcriptome to the proteome is further enhanced by a process called protein post-translational modifications (PTMs).

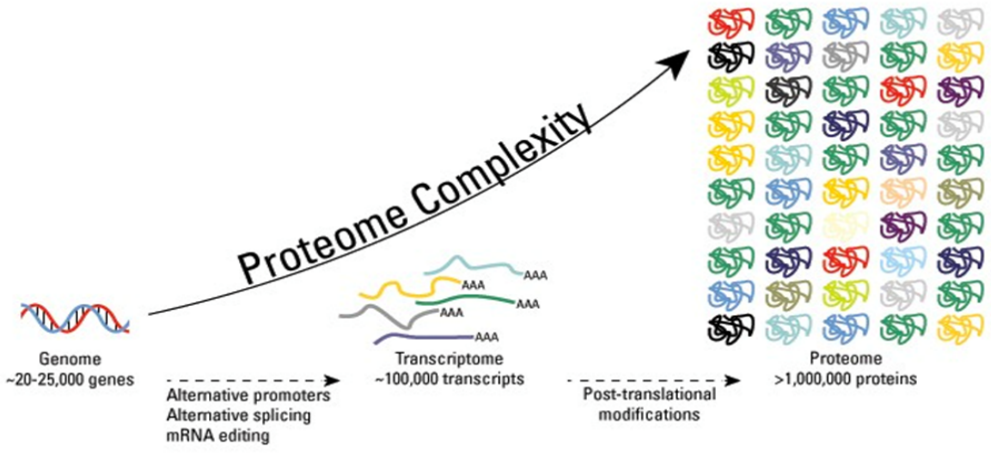


Figure 1.1. Post-translational modifications are important mechanisms to expand the diversity of proteome. The genome is consisted of 20,000-25,000 genes, yet the proteome is estimated at over 1 million proteins. The size of transcriptome is increased by the changes at the transcriptional and mRNA levels, and the myriad of different post-translational modifications dramatically increases the complexity of the proteome compared to both the transcriptome and genome [1].

PTMs refer to covalent modifications that occur after DNA has been transcribed into mRNA and then translated into proteins. The nascent or folded proteins, which are stable under certain physiological conditions, are then subjected to a machinery of specific enzyme-catalyzed modifications on their side chains or backbone [3]. Proteome diversification by covalent modification occurs in prokaryotes but is much more extensively encountered in nucleated cells, both in terms of types of modifications and frequency of occurrence [4]. PTMs involve the covalent addition of a functional group to a protein, where the functional group includes phosphorylation, hydroxylation, glycosylation, methylation, ubiquitination, acetylation, and sulfation as shown in Fig. 1.2 [5]. PTMs serve to functionally diversity the proteome by finely tuning the structure, stability, activity, subcellular localization, and protein interaction partners of the modified proteins [3-5].

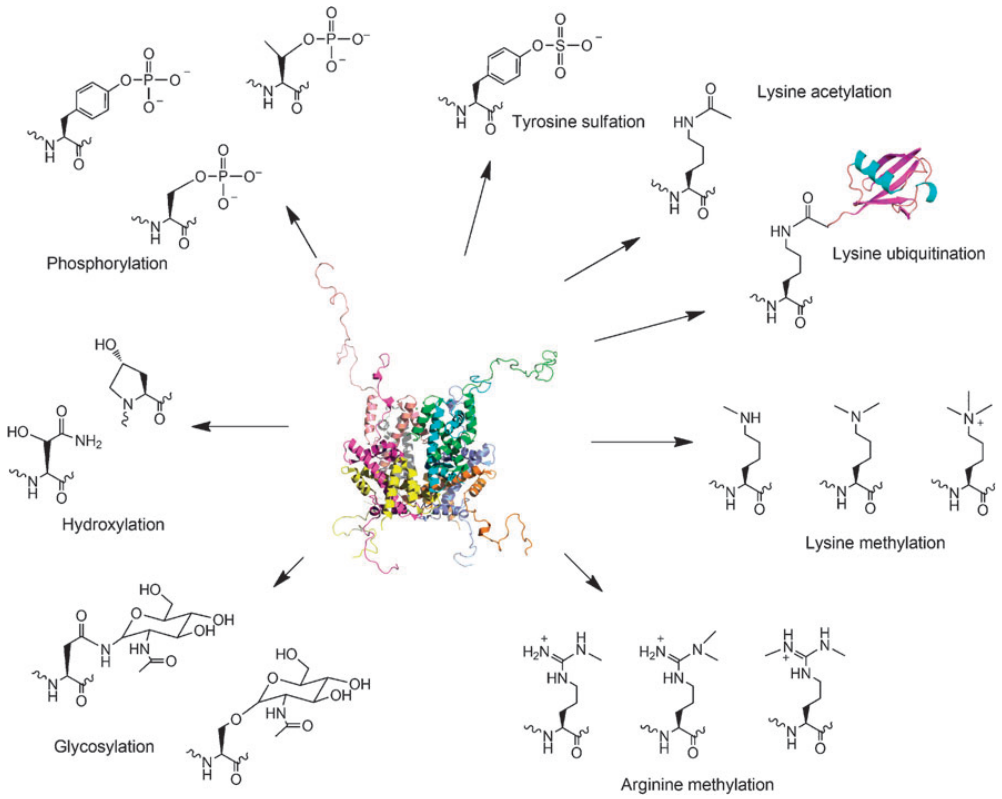


Figure 1.2. Covalent modification to protein side chains. Phosphorylation, hydroxylation, glycosylation, methylation, ubiquitination, ubiquitination, acetylation, and sulfation [5].

Of all the post-translational modifications, phosphorylation and glycosylation are the major players in many of the protein functions. Based on quantification of post-translational modifications levels, phosphorylation dominates the number of experimental PTMs identified by an order of magnitude, while N-linked glycosylation dominates the number of putative PTMs as depicted in Fig 1.3 [6]. Indeed, phosphorylation easily ranks as one of the most common PTMs in eukaryotes with more than 100,000 phosphosites experimentally confirmed in humans and/or related mammals and over two-thirds of the 23,000 proteins encoded by the human genome demonstrated to be covalently modified with phosphate by the collective activity of

~500-1000 protein kinases [7-10]. Phosphorylation is particularly quintessential in signal propagation where it regulates the function of numerous proteins in signaling networks by activating or inhibiting enzyme activities [11-13]. Phosphorylation of proteins can regulate protein function in two main ways. First, phosphorylation at specific residues may induce changes in protein conformation, thus leading to altered function such as enzymatic activity or ion channel permeability. Alternatively or in parallel, phosphorylation can introduce specific binding sites for other proteins, enabling the assembly of functional complexes of proteins [14]. Phosphorylation of transcription factors is crucial to gene regulation while phosphorylation of cytoskeletal proteins is essential for control of cell shape and motility. Lastly, recent advances indicate that phosphorylation is an initial step in targeted protein degradation [15]. Considering the pivotal role of phosphorylation in signal transduction, it is not surprising that malfunction of signal transduction pathways through phosphorylation (or lack thereof) either directly causes or is a consequence of many human diseases, such as cancer and neurodegenerative disorders [16].

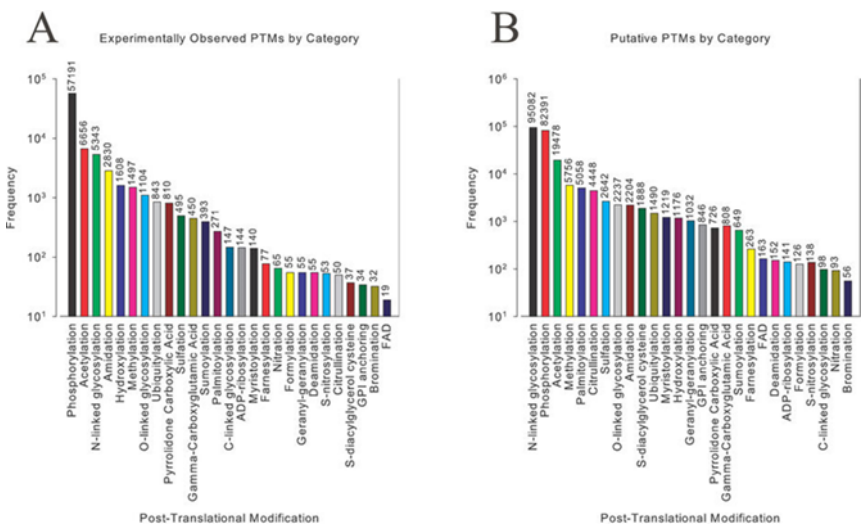


Figure 1.3. Summary of (A) Experimental and (B) Putative Post-Translational Modifications Curated from Swiss-Prot [6].

Over the last two decades, mass spectrometry-based proteomics has emerged as one of the most effective approaches for analyzing PTMs and identifying their sites of attachment on proteins, including phosphoproteins [17-20]. Given the steady increase in the number of functionally important phosphorylation sites that have been uncovered, there is a growing need for phospho-specific binding molecules [21] that can be developed for traditional biochemical approaches as well as advanced techniques such as single-cell analysis [22-24] and high-throughput assay systems [25-27]. The most common affinity reagents for detecting PTMs, and more specifically phospho-modified sites, are conventional monoclonal antibodies (mAbs) that have been raised in mice [28]. In addition to full length antibody, in particular IgG, other formats of antibody based binding scaffolds, for example, a single-chain fragment variable (scFv) and a single variable immunoglobulin domain denoted VHH (camelids) or V-NAR (cartilaginous fishes) could be used for phosphor-specific binders. Apart from antibody

based binding scaffolds aforementioned, alternative binding scaffolds, including anticalins, affibody molecules, adnectins, and DARPins, are able to be potential candidates for PTMs specific binders. One of the most promising alternative protein scaffold for phospho-modified targets was DARPin on which this thesis is based. Hereunder are a review of antibody binding scaffolds and alternative protein scaffolds.

2. Protein scaffold for molecular recognition

In the final decade of the 19th century, Emil von Behring and Shibasaburo Kitasato developed serum therapy for the treatment of diphtheria and tetanus [29]. This finding was awarded the first Nobel Prize in Physiology or Medicine in 1901 to Behring for his work on serum therapy, in particular its application against diphtheria. In 1897, in order to explain the process of immunity, Ehrlich postulated that the side-chain model of immunity to explicate his experimental observations. He had a hypothesis that injurious compounds (toxins) could mimic nutrients for which cells express side chains that he called 'nutriceptors' [30-31]. He described the notion of the magic bullet by proposing that invading microbes could be specifically targeted without damage to the host. In 1908, he was jointly awarded the Nobel Prize in physiology or medicine, with Élie Metchnikoff, in recognition of his contributions to understanding immunity. Today it is well known that nature's solution for the magic bullet with high affinity are antibodies or immunogloblulins which are proteins consisting of several polypeptide chains that are able to bind different antigens ranging from small molecules called haptens to large macromolecules. Conventionally, antibody production can be accomplished by animal immunization, which can generate either polyclonal antibodies

or monoclonal antibodies. Monoclonal antibody production was first pioneered by Milstein and Köhler who were later awarded Nobel prize in 1975. In principle of monoclonal antibody generation, a laboratory mouse is immunized by an antigen. Antibody producing cells called plasma cells (active B cells) are then harvested and fused with immortal cancer B cells (myeloma cells) to obtain immortal hybrid cell lines (hybridomas) as a renewable source for monoclonal antibody production. With the advancement of recombinant DNA technology, engineered antibodies in different formats, including scFv or VHH were generated, and these engineered antibodies are able to be produced at high yields in microbial cells, including *E. coli*. Apart from antibody binding scaffolds, the concept of binding agents for molecular recognition has been further extended to non-antibody based or alternative protein scaffolds, including anticalins, affibodies, adnectins, and DARPins that this thesis is focused on.

2.1 Natural polyclonal or monoclonal antibodies

Antibodies or immunoglobulins are an important protein in an adaptive immune system of human and higher vertebrates. In order to protect body from foreign antigens, including parasites, bacteria, viruses, fungi, pollens, or self-antigen like cancer cells, specialized white blood cells denoted plasma cells secrete antibodies into body fluids. The natural function of antibodies is to bind selectively to antigens, including proteins, nucleic acids, carbohydrates, or haptens that activate humoral immune response. Antibodies protect the host by a functional neutralization of antigen after their binding. For example, antibodies can neutralize viral infectivity by interfering with virion binding to receptors, block uptake into cells, prevent uncoating of the genomes in

endosomes, or cause aggregation of virus particles [32]. After binding to targets, a process of elimination of bound antigen be achieved by the activation of a cascade of complement proteins resulting in pore formation and cell lysis.

Immunoglobulins are generally made up of two identical heavy and two identical light polypeptide chains connected with disulfide bonds. The light chains are consisted of one variable domain VL and one constant domain CL while the heavy chains are consisted of one variable domain VH and three to four constant domains which are CH1, CH2, CH3 and CH4. All these domains are of a typical immunoglobulin domain fold of two stacked β-sheets that are stabilized with conserved disulfide bonds. Igs are classified according to their heavy chain type (α/IgA, δ/IgD, ε/IgE, γ/IgG, μ/IgM) and light chain type (λ, κ) [33].

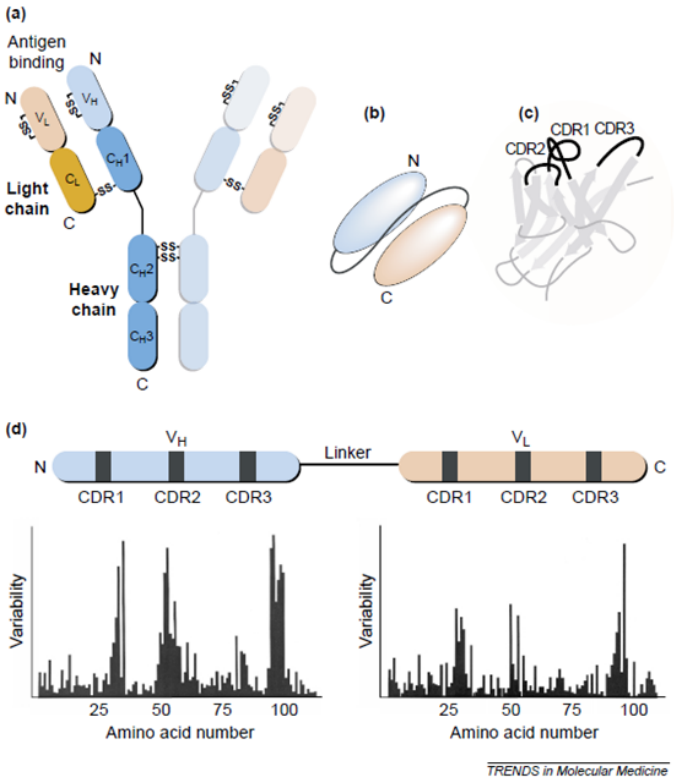


Figure 1.4. Schematic illustration of an IgG molecule and a single-chain fragment variable (scFv) IgG derivative. (a) IgG is composed of variable domains (V) and constant domains (C) on heavy chain (H) and light chain (L). Both heavy and light chain domains contain intradisulphide-bonds (S-S) and heavy and light chains are also interconnected with disulfide bonds. (b) and (c) scFv is composed of variable domains of heavy and light chain (VH, VL) joined with a peptide linker. (Lobato & Rabbitts, Intracellular antibodies) [34].

The antigen binding site is located on the tips of the Y-shaped antibody molecules and both in VL and VH contribute with three variable loops each called complementary determining regions (CDR1/2/3) that form a continuous and very plastic binding surface [34] as illustrated in Fig. 1.4. The stunning variability and plasticity of the antigen binding site is probably the reason for its ability to bind a manifold of different antigens with high affinities. These adaptive and versatile binding features are caused by a high sequence diversity of the six variable loops, several possible CDR loop conformations, fine-tuning of shape complementarity by single somatic mutations in the CDR framework [35] and lastly the possibility of an induced fit mechanism that enables structural adaptation to the antigen [36].

The debut of the production of phosphorylation-dependent antibodies was reported in 1981. The technology used to produce these antibodies was immunization of rabbits with benzyl phosphonate conjugated to keyhole limpet hemocyanin (KLH), which generated polyclonal antibodies that recognized phosphotyrosine-containing proteins [14]. These antibodies became important reagents used in oncogenic virus and cancer research; however, their utility was limited success because they bound to any protein containing the phosphorylated amino acid. As a result, they were most useful in Western blotting applications, in which multiple phosphoproteins could be resolved by their molecular weight. Researchers have also applied these antibodies in

immunohistochemical studies of human cancers and showed elevated accumulation of phosphotyrosyl-containing proteins in cancers [37]. Shortly thereafter, there was another report describing the production of serum antibodies that differentiated between the phospho- and dephospho-forms of G-substrate, a protein that resides in cerebellar Purkinje cells and is phosphorylated by cGMP-dependent protein kinase [15]. The authors used a synthetic heptapeptide, Arg-Lys-Asp-Thr-Pro-Ala-Leu, as antigen. The synthetic peptide corresponds to a repeated sequence surrounding two phosphorylated threonyl residues in the intact protein. Phospho-specific antibodies were generated by rabbit immunization with the purified phosphoprotein, which was prepared by *in vitro* phosphorylation with cGMP-dependent protein kinase. Despite these successes, many other attempts to produce phospho-specific polyclonal antisera by immunization with the phospho-form of intact proteins have not been successful, probably because of two significant factors. First, many phosphorylated proteins are believed to undergo rapid dephosphorylation during immunization, regardless of the route of injection, leading to the loss of the desired phospho-epitope. Second, holoproteins generally contain multiple immunogenic epitopes. This decreases the probability that clonal dominance for a phospho-specific epitope will be obtained. Nevertheless, the generation of a polyclonal phosphospecific antibody is often imprecise, low-throughput, expensive, time-consuming and not renewable. Related to that, the production of monoclonal phosphospecific antibodies needs additional screening of copious hybridomas, which is made more challenging by the scarcity of phosphospecific antibody clones, around 0.1–5% [38-40]. Ultimately, disproportionately more phosphotyrosine (pTyr)-specific antibodies exist than phosphoserine (pSer)- or phosphothreonine (pThr)-specific

antibodies. As a result, this has hampered the study of serine and threonine phosphorylation, which comprises 90% and 10% of all phosphorylation sites, respectively, compared with less than 0.05% for tyrosine [41]. Therefore, for most targets, no specific reagents exist, and in cases where commercially mAbs are available, they are known to be of highly variable quality and limited utility [42-43].

2.2 Recombinant generation of antibody fragments and monoclonal antibodies

As an alternative to animal immunization and hybridoma technology, recombinant DNA technology enables the generation of engineered antibody fragments and full-length monoclonal antibodies with desired traits from natural or synthetic library sources. This alternative was established during progress in the field of antibody fragment cloning, engineering and production as well as protein selection technology [33].

The productions of three different functional antibody fragment formats in *E. coli* were published in Science in 1988. Firstly, non-covalently associated VL and VH was generated by Skerra and Plückthun [44]. Secondly, VL carboxyl terminus was covalently attached with a designed peptide flexible linker to the VH amino terminus, yielding the single-chain fragment variable or scFv, Fig. 1.4 (b) and 1.4 (c) by Bird and colleagues [45]. Finally, a complete Fab fragment was produced by Better and colleagues [46]. Both the scFv and the Fab fragment has proved to be versatile formats for protein engineering because they can be further engineered or modified by recombinant DNA technology. Importantly, they can be produced at high yields in

microbial hosts which are much cheaper compared to mammalian cells, and they could be combined to generate bispecific or trispecific variants [47] as shown in Fig. 1.5.

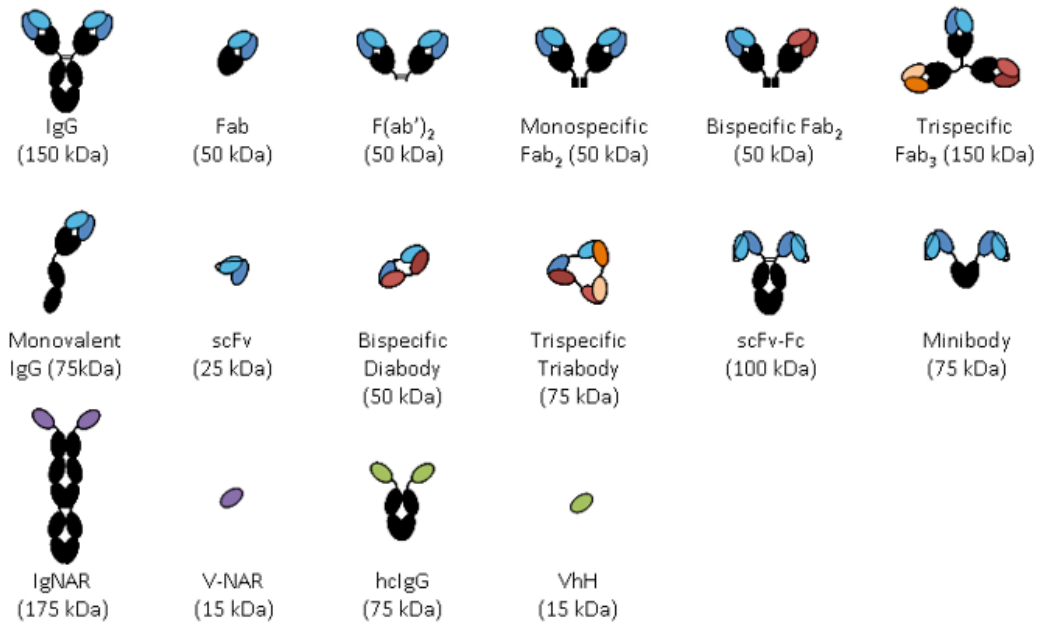


Figure 1.5. A schematic representation of different antibody formats. Classical IgG molecule, camelid heavy chain IgG (hcIgG) and shark IgNAR alongside antibody fragments generated from these [47].

Principally, antibody fragments from *in vitro* generation with desired binding affinity and selectivity involve the construction of a library pool encoding the variable VL and VH antibody domains to be displayed on a protein selection platform, including phage, ribosome or cell display. Both scFv and Fab are two most frequently used for surface display. Libraries of displayed fragments are then proceeded to a biopanning or screening procedure as described in phage display section. Sources of library pools can be immune or naïve natural repertoires as well as synthetic ones. Compared to animal immunization and hybridoma technology, *in vitro display* selection provides the protein

engineer more set screws, e.g. to assure a functional representation of the antigen, direct the selection towards particular epitopes or select for other traits than affinity [33-35]. Engineered antibody fragments provide more advantages than full-length antibodies because they are (1) much smaller and monovalent fragments so they are potentially used for cancer imaging and radiotherapy applications, (2) more efficiently penetrate the tumor and are faster clearance from the circulation, i.e. reducing radiotoxic side effects [47-48]. In the event that full-length antibody is required, *in vitro* selected antibody fragments can be converted into a full-length antibody format and produced in eukaryote cell culture.

2.3 Single domain antibody fragments

In addition to IgG which is human-derived antibodies, other animals, such as camelids (camels, and llamas) and cartilaginous fishes (nurse sharks) could produce other formats of antibodies [49]. In both animals, the variable antigen binding entity is consisted of a single variable immunoglobulin domain named VHH for camelids) or V-NAR for cartilaginous fishes. The VH of "classical" antibodies has a relatively hydrophobic face, usually associated with the light chain, which makes it aggregation-prone in solution when produced alone [47, 49]. With the help of recombinant DNA technology, stable genetically engineered single VHH domains with a MW of approximately 15 kDa can be produced in microbial hosts, including yeast or *E. coli*. Some VHH antibody fragments contain an intramolecular disulfide bond between CDR1 and CDR3. The intramolecular disulfide bond might play a role in shaping the loop structure, and may require proper protein folding in an oxidizing environment for

stability [49]. Similar to antibody fragments, VHHs with desired traits can be obtained from synthetic, naïve or immune libraries that are displayed on protein selection system, including phage, cell, or ribosome display and selected for antigen binding. VHHs' applications included *in vitro* diagnostics, nanosensors as well as *in vivo* cancer imaging, the treatment of rheumatoid arthritis, Crohn disease or thrombosis [47]. However, their immunogenicity may be a concern for therapeutic applications because VHHs are not stemmed from a human protein scaffold.

2.4 Alternate non-antibody binding scaffolds

From the lessons learned with antibody-derived binding scaffolds, they all contain disulfide bonds which are necessary for proper folding which involves stability and affinity; therefore, it is difficult for intracellular application, especially phospho-modified targets which are typically intracellular located. In an attempt to generate smaller binding scaffolds with high affinity binding and stability independent of disulfide bonds, many synthetic biology tools were developed to enable the creation of gene libraries and selection techniques, including phage display, ribosome display, surface display, protein complementation, etc. [50]. However, with these synthetic technologies the use of the immunoglobulin scaffold itself became dispensable and thus protein engineers began to utilize novel binding proteins with the goal of improved biophysical properties, including stability, solubility, multi-valency and modification capabilities [51].

Most of the proteins used as scaffolds are naturally involved in protein binding although they show a large diversity in structure and function. The choice of binding

scaffold proteins is typically dependent upon the intended future purpose of the generated affinity ligands; however, some generally important features of binding scaffolds can be formulated [51-52]. First, the absence of aggregation is not only favorable for an efficient selection process but it is also mandatory for practically all in vitro applications — and absolutely mission-critical for in vivo applications. Second, many applications require chemical coupling, for example, to fluorophores, toxins, nanoparticles or solid supports. This is much facilitated if the protein can be engineered to have a unique cysteine. Third, the absence of disulfide bonds is a prerequisite for high yield and functional expression in bacterial cytoplasm; proper folding in the cytoplasm of higher cells will also allow functional studies. Fourth, multivalence and multispecificity are very generic concepts to increase avidity, or to crosslink different epitopes or targets. Multimeric assemblies should ideally have the same expression yield as a monomer. Furthermore, different linking geometries should be realized with ease. Fifth, scaffolds should be able to give rise to well expressing fusion proteins that do not lead to additional aggregation. Sixth, scaffolds should allow convenient engineering for in vivo half-life extension; for example, by site-specific PEGylation, or by fusion to a domain that binds to serum proteins with long half-lives themselves.

The protein should be stable enough to be highly tolerant to modification by randomization so that a library with functional members can be generated. A library of a chosen protein scaffold is generally created by selective random mutagenesis of an appropriate number of surface exposed residues. The origin of the protein used as a scaffold should be also taken into account. If the affinity proteins are designed for therapeutic purpose, the problem of potential immunogenicity should be an issue to

consider. If the protein is of foreign origin it will be likely to cause some immune response if no precaution is taken, but also scaffolds based on human proteins would have the potential of becoming immunogenic by the introduced amino acids and altered binding sites. In addition, when using human protein scaffolds, the risk of causing autoimmunity reaction would, at least in theory, need to be considered.

The different binding scaffolds are most commonly classified based on their structure and the utilized binding-site engineering strategies. The field of alternative protein scaffolds currently shows a trend of consolidation towards a few classes that are tested extensively in different applications including clinical trials [35]. Some of them, including anticalins, affibody, adnectins, and DARPins (Fig. 1.6) that this thesis is based upon, will be presented here in more detail.

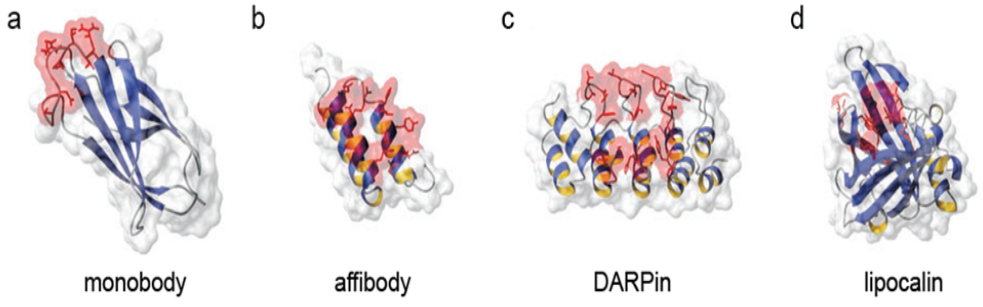


Figure 1.6. Alternative scaffold proteins. Each scaffold protein domain is shown as a ribbon diagram with residues randomized for library creation highlighted in red. (a) The monobody structure show loop diversity similar to scFv domains, while (b) the affibody utilizes a flat surface for binding. (c) The DARPin repeat scaffold utilizes both surface and loop diversity in library creation, while (d) lipocalins have a recessed cavity for binding smaller molecules. Images taken from Binz et al. [53]

2.4.1 Anticalins

Anticalins are a class of engineered ligand-binding proteins that are derived from the lipocalin scaffold, a family of proteins found in a variety of organisms, such as bacteria, plants, insects or vertebrates [54]. Natural lipocalins occur in human plasma and body fluids with the nature function to bind, store or transport vitamins, steroids or metabolic compounds [54-55]. The lipocalin protein architecture is characterised by a compact, rigid β-barrel fold that supports four structurally hypervariable loops which form a pocket for the specific complexation of differing target molecules. The β-barrel fold consists of eight anti-parallel β-strands winding around a central axis, and typically a C-terminal α-helix that is associated to the β-barrel.

Lipocalins are about 21 kDa in size, and also typically contain one or two disulfide bonds, as commonly found in proteins secreted from eukaryotes [54-56]. The β-barrel fold is highly conserved, whereas the ligand entry site including four loops connecting the β-strands is varied among different lipocalins. In earlier studies, Skerra and colleagues used the bilin-binding protein (BBP) from *Pieris brassicacea* as a scaffold for the randomization of 16 positions located within the four loops or adjacent β-strands at the ligand entry side [57]. With this library selection, they isolated binders with nM affinities for the haptens fluorescein and digoxigenin by phage display [57]. Apart from fluorescein and digoxigenin, BBP variants were so far selected against several other low-molecular weight substances and peptides, including phthalic acid esters [58], thus showing the suitability of the lipocalin scaffold for the generation of novel ligand-binding proteins. Later, anticalins were also extended to select for binding protein targets. The cytotoxic T lymphocyte-associated antigen (CTLA)-4 was chosen

as a target with an impact in oncology for the development of an antagonistic anticalin based on the NGAL scaffold [57]. To generate CTLA-4 specific anticalins, a combinatorial phagemid library was generated by randomising 20 amino acid positions at the tips of the four loops in the three-dimensional structure of NGAL [57, 59]. By combining the phagemid library with phage display selection system, anticalin variants with affinities in 10 nM region were isolated.

Since most anticalins contain disulphide bonds, they require an oxidizing environment for proper folding. The secretion in the oxidizing *E. coli* periplasm was described and a co-secretion of proteins that support the formation and isomerization of disulfide bonds may improve the yield of functional protein [60]. However, due to the need of disulfide bonds, this limits the potential intracellular application where cytoplasm is a reducing environment.

2.4.2 Affibody

In 1997, Nord and colleagues [61] demonstrated that a small scaffold with three-helical subdomain could be used as a general scaffold for diversity generation. This scaffold is derived from the B domain of the *Staphylococcus aureus* cell surface receptor protein A. The function of protein A is to bind to Fc-portion of immunoglobulins produced from host serum and thereby evading and modulating host immune system [62]. All of the five highly homologous immunoglobulin-binding domains of protein A, denoted E, D, A, B and C, share a three-helix bundle fold. The B-domain is a relatively short cysteine-free peptide of 58 amino acids that is folded into a three-helical bundle structure and the kinetics of the folding reaction is one of the

fastest that has been reported [63]. The folding properties in combination with high solubility and a relatively high thermal stability has contributed to a high scientific interest. The B-domain was first mutated at key positions mainly for enhanced chemical stability by exchange of glycine for alanine at position 29 for an increased resistance towards hydroxylamine, and by domain end substitution to facilitate head-to-tail dimerization at DNA level [64]. The resulting engineered variant was called the Z-domain [64-65]. Compared to B domain binding to both Fab and Fc fragments of IgG in a species, the engineered Z-domain retained its affinity for the Fc part of the antibody while the weaker affinity for the Fab region was almost completely lost.

With only 58 amino acids and a molecular mass of about 6 kDa, the Z domain is only half of the size of an antibody variable domain (V_H or V_L) but still provides a binding interface of about 800-900 $Å^2$, that is typically found in protein interfaces [65]. The Z domain folds rapidly and does not contain any cysteine residues for disulfide bond formation, and thereby allowing for the production and folding in reducing environment, including the *E. coli* cytoplasm [66].

Previous libraries were typically created using NN(G/T) or (C/A/G)NN degenerate codons, while the latest libraries are built of primers with defined sets of trinucleotide codons and are based on a further engineered scaffold for higher thermal stability, higher hydrophilicity and more efficient peptide synthesis of library members [67]. Selection systems used in combination with library construction includes predominantly phage display, phage in combination with *Staphylococcus* cell surface display [68], and also ribosome display [69]. Selected binders have been obtained against various targets, including *Taq* DNA polymerase [70], transferrin [71], TNF-α

[72], Her2 [73], IGF-1R [74], other affibody molecules [75-76], IgA [77], HIV gp120 [78], CD28 [79] but also the amyloid beta peptide [80].

Because of its small size, high affinity and promising pharmacokinetics such as tumor penetration and fast renal clearance, the affibody molecule Z_{HER2} *was* proved to be a suitable affinity reagent for the radioimaging of Her2-expressing tumors in a mouse xenograft model [81] and has already entered clinical studies in humans for *in vivo* imaging of breast cancer [83]. Apart from detection purpose, therapeutic affibody applications have been investigated by Rafal and colleagues; a truncated format of the *Pseudomonas* exotoxin A (PE38) was fused to Z_{HER2} for a selective delivery of the toxin to Her2-expressing cancer cells, and they found a correlation between Her2 expression and PE38-mediated inhibition of protein synthesis [84].

2.4.3 Adnectins

Another interesting class of alternative protein scaffold are adnectins, a family of designed proteins based on the framework of the 10th human fibronectin type III domain (^{10}Fn3). The 10th fibronectin type III domain is used as the starting template for the design of a family of target-binding proteins due to its structural similarity to antibody variable domains, suitability for modular assembly into multi-functional molecules, favorable biophysical properties and its abundance in human blood and extracellular matrix, which demonstrates that inherently this scaffold is not toxic or immunogenic [85]. Despite of the lack of significant sequence homology between antibody variable domains and ^{10}Fn3, they share similar structures as illustrated in Fig. 1.5. Both antibody variable domains and ^{10}Fn3 consist of sandwiches of two anti-

parallel beta sheets, with solvent-accessible loops at each pole of the domain that resemble an antibody' CDRs. Unlike immunoglobulin-domains, the ^{10}Fn3 does not contain any disulfide bonds and thus retains its stability under reducing environment [86-87]. These parameters make it a very suitable scaffold for the introduction of novel binding selectivities.

In earlier study, the sequences of 17 naturally occurring ^{10}Fn3 domains were aligned and decided to variegate the variable, adjacent BC and FG loops with five randomized aa positions, respectively [88]. The randomized library of 10^8 clones was displayed on phage and productive binders against ubiquitin could be selected [88]. In most of the selection experiments performed to date, either two or three loops located on one of the poles (BC, FG or BG, DE, FG loops) were diversified and binders with μM to nM affinities have been selected. Applied selection systems for adnectins reported to date include phage, yeast and mRNA display. Combination of error-prone PCR and DNA shuffling procedures with these display technologies, clones with pM affinities were selected [89]. The reported antigens are to date exclusively proteins, including VEGF-R2. An adnectin selected against VEGF-R2 was PEGylated for an extension of if half-life in the circulation and has already entered clinical phase II trials for the therapy of solid tumors [85].

2.4.4 DARPins

A suitable scaffold will have to replicate the virtues of antibodies and address their shortcomings [90]. Among alternative binding scaffolds, repeat proteins are very attractive as an alternative to be a general binding protein. They are typically comprised

of repeat modules stacking on each other to form a compact folded domain, usually with an elongated shape [22], and use variable surface residues to create an extended target interaction interface. The conserved interfaces between the repeat units allow individual repeats to be further engineered, including exchange, deletion, or insertion without destroying the tertiary structure of the domain [90].

Pancer and colleagues [91] reported the surprising discovery that the immune system of jawless fishes that may represent an evolutionary remaining of early vertebrates employs leucine-rich repeats instead of immunoglobulins for the generation of antigen-specific binding proteins [92]. Several repeat proteins such as the Ankyrin, WD, leucine-rich or tricopeptide repeats have been subjected to protein engineering; however, this section focuses on only ankyrin repeat proteins since this thesis is based upon and they progress the furthest than all others toward biomedical applications.

The ankyrin repeat (AR) proteins are natural binding molecules and have been found across all phyla ranging from viruses to humans. The ankyrin repeat has been discovered in more than 400 proteins such as cell cycle regulators, cytoskeletal organizers and tumor suppressors [93-94]. These proteins are present in nucleus and cytoplasm indicating that they can function in reducing condition, and adapt to different environments. Ankyrin repeat domains (AR) are repeating structural units of 33 amino acids, which stack together to form elongated proteins [95]. In terms of architecture, each repeat consists of a beta turn followed by two antiparallel alpha helices and a loop reaching the turn of the next repeat. The number of repeats within one protein is variable.

To take advantage of the properties of ankyrin repeat proteins as binding molecules, a library was engineered by a method called consensus design [96]. In this

strategy, six of the 33 amino acids in each consensus repeat were randomized and libraries containing 2 or 3 repeats were generated. To shield the hydrophobic core a capping repeat was added at each the N and C terminus (Fig. 1.7). Members of the libraries are denoted Designed Ankyrin Repeat Proteins (DARPins). Not only did this strategy produce large libraries but also DARPins with optimal biophysical properties [97]. From this DARPin design, these libraries are a valuable source of binding molecules, which was shown by selection of high affinity binders against several types of proteins, including MBP, p38, Her2, JNK2, APH, ERK2, and pERK2 selected using either ribosome or phage display [53, 98-99]. The simple molecular and modular architecture of DARPins allows a broad range of applications. This, taken together with their high affinity and selectivity makes DARPins attractive candidates for further drug development.

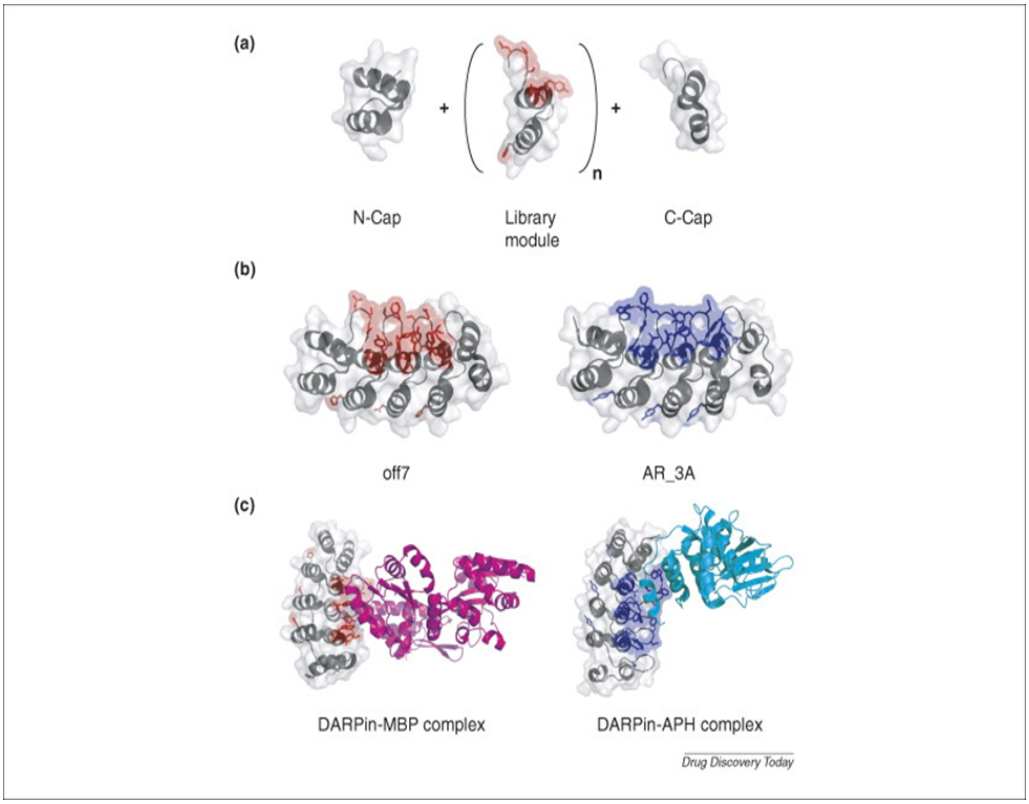


Figure 1.7. Schematic representation of the DARPin library, the resulting DARPin molecules and the target protein binding. (a) DARPin libraries are composed of 3 elements, including N-Cap, C-cap, and a variable number of internal repeats (*n*) which is typically between 2-4 repeats, giving DARPins' size between 14 and 21 kDa, respectively (b) The three dimensional structure of two DARPins, Off7 and AR_3A, both with three library modules (n=3), is shown. The blue and red color represent the randomized surface for potential target interaction. **(c)** Two DARPins (Off7 and AR_3A) with their target protein are shown (MBP, maltose binding protein and DARPin off7 PDB: 1SVX; APH aminoglycoside phosphotransferase and DARPin AR_3A, PDB: 2BKK). [52]

3. Protein selection technologies

As discussed in the previous section, alternative binding scaffolds, including anticalins, affibody, adnectins, and DARPins could be promising binders for either immunodetection or immunotherapy. These could be achieved by combination of these scaffolds, recombinant DNA technology, and protein selection technologies. The

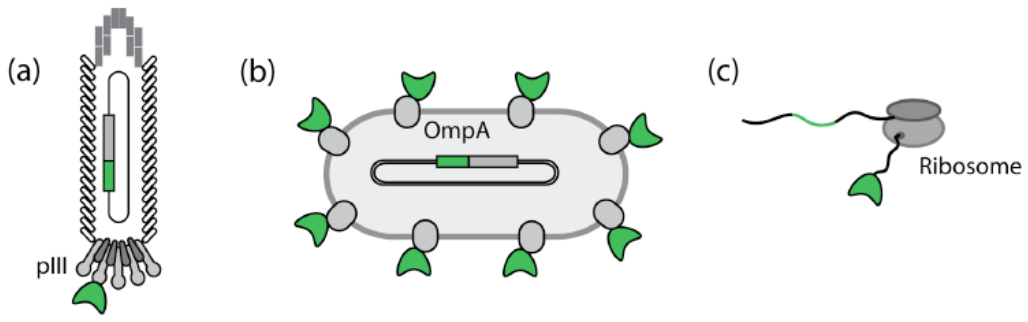
focus of this section will be on systems used for a functional selection of variants with improved properties among those populations.

In natural selection, certain traits within a diverse population are more ubiquitous in subsequent generations since they contribute individuals a fitness advantage, particularly in certain environmental stress conditions, including predators, osmotic stress, heat, or coldness [35]. These traits confer element of the individuals' phenotype or observable characteristics, such as their size, shape or color. Each phenotype has its own matching genotype, the genetic blueprint that is passed on to next generation and ensures the heritability and propagation of the phenotype. In the laboratory, protein selection platforms imitate three fundamental features of natural selection: (1) a diverse starting repertoire, (2) a coupling between genotype and phenotype, i.e. a linkage of a protein with its encoding nucleic acid is established that allows for both an identification of the selected protein variants by recombinant DNA technology and a continued diversification and selection process, (3) a selection pressure that confers the protein variants with the desired traits, i.e. a functional selection of the fittest, including catalysis, stability, binding affinity and selectivity.

A powerful protein selection experiment could be varied depending on what platform is used; however, some general requirements have to be met. The conditions should be optimized in order that only the desired protein trait is selected for, and the backgrounds owing to protein expression levels, microbial growth or avidity effects are eliminated. The enrichment should be efficient. For instance, some proteins as full-length antibody, antibody fragments, or some binding scaffolds may require an

oxidizing environment for disulfide bonds formation to obtain their proper folding, and function.

In addition to selection pressure, the size of a library does have a strong influence on the consequence of a selection. Absolutely, larger size of functional libraries will provide more sequence and structure space and thus potentially yielding desired traits. Finally, a fast selection technology is certainly desirable since post-selection work is usually more time-consuming than the actual selection process. Protein selection technologies commonly used in protein engineering research (Fig. 1.8) include phage display, cell display, ribosome display, mRNA display, CIS display, and protein fragment complementation assay that will be covered in sections below. In addition to these existing technologies, a new technology denoted FLI-TRAP (Functional Ligand-binding Identification by Tat based Recognition of Associating Proteins) that this thesis is based upon will be detailed.



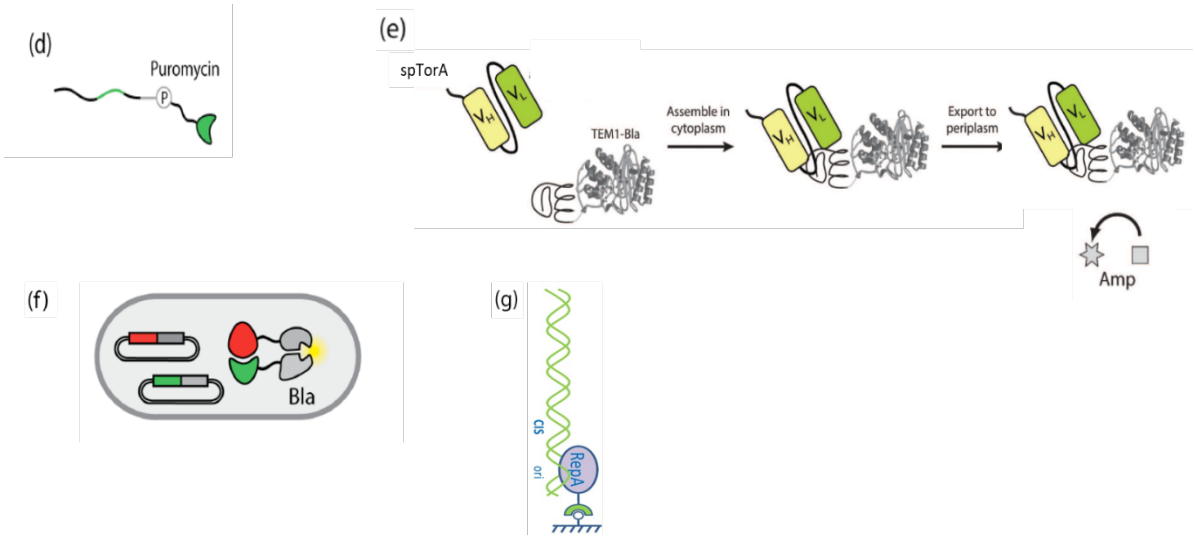


Figure 1.8. Schematic overview of different protein selection systems. The foreign proteins and their cognate genes are indicated in green. Systems (a-d) are display systems while systems (g) and (h) are compartment systems. (a) phage display: the foreign protein is fused to the M13 phage pIII and incorporated in the phage coat; (b) cell surface display: exemplarily, OmpA anchors the foreign protein to the cell surface; (c) ribosome display: a stalled ribosome attaches the foreign protein to its cognate mRNA; (d) mRNA display: puromycin links the foreign protein covalently to its cognate mRNA; (e) FLI-TRAP: spTorA is fused to scFv while target protein is fuse to TEM1-Bla; colocalization of protein complex confers antibiotic degradation, and rendering cell growth; (f) protein fragment complementation assay: the binding between bait and scavenger reconstitutes the function of two halves of β-lactamase (Bla); (g) CIS display: RepA-protein binds to immobilized target, and RepA binds to its coding DNA sequence. Figure is adapted from [33].

3.1 Phage display

Phage display is a protein selection platform that primarily involves utilizing filamentous phage [100]. Phage display involves the generation of large combinatorial libraries of peptide, protein, antibody, or other binding scaffolds displayed on the surface of a phage particles that encode the sequence of the displayed protein. Protein display on virion surface was first reported by Smith in 1985 [101]. In this work, foreign DNA fragments from EcoRI restriction endonuclease were fused to the gene for protein III in f1 filamentous phage DNA. The engineered fusion proteins could be

incorporated into correctly assembled virion coats, and displayed on the virion surface. The displaying phages still had their ability to infect *E. coli* and could pass their genetic information to progenitor phages. His finding opened a gate for library based protein engineering research. Later, phage display was successfully utilized to display human growth hormone [102] and the first antibody fragments [103]. In 1991, antibody fragments binding to an antigen could be selected from a phage display library [104]. Following years, desired protein traits, including affinity, stability, and catalysis were selected on phage [105]. Selections are typically performed *in vitro*, but sometimes even in higher organisms [106-107].

The filamentous bacteriophages F1, fd and M13 are a group of single-stranded DNA containing viruses with the ability to infect *E. coli* cells containing the F conjugative pilus as a receptor [33]. Among filamentous bacteriophages, the most commonly used phage for display is M13 which. Its single stranded, covalently closed genome confines the length of the wild-type phage particle and is encapsulated by around 2700 copies of the major coat protein pVIII [33]. The ends of the phage are decorated with about 5 copies of pVII and pIX on one end and 5 copies of pIII and pVI on the other (Fig. 1.9). Using phage vectors with single copies of coat protein genes, foreign proteins are commonly displayed on pIII [108-109], while short peptides can also be displayed on pVIII [110-111].

To be efficiently displayed on surface, the foreign proteins need to be efficiently translocated to the bacterial envelope or across membrane to periplasmic space. The commonly used translocation patways include secretory (Sec) pathway which

translocates proteins in their unfolded state. However, some proteins which are fast-folding and stable proteins such as thioredoxin or DARPins are not efficiently translocated by the Sec pathway. An alternative system called the signal recognition particle (SRP)-dependent pathway could be employed. This patway translocates the nascent polypeptide chain during translation and was proposed for display of very stable and fast-folding proteins and scFv antibody fragments [112-113]. A third system is the twin-arginine translocation (Tat)-mediated pathway that translocates native, folded proteins that fold in the *E. coli* cytoplasm [114]. These three systems can be performed by simply choosing an appropriate signal sequence.

A typical protein selection experiment employing phage display starts with a library gene pool that is enzymatically ligated with linearized phage or phagemid DNA and then circularized vectors are transformed into competent *E. coli* cells. A schematic of phage display selection cycle for isolation of binders to an antigen of interest is illustrated in Fig. 1.10. A pool of phages is exposed to the antigen either immobilized on a solid phase such as a paramagnetic bead or microtiter plate (solid phase), or in solution followed by post-incubation capture onto a solid support (solution phase). Unbound or weakly-bound phages are removed by repeatedly washing the solid support. The selection condition can be manipulated by adjusting the concentration of antigen, the number and duration of washing steps or the concentration of a competitor antigen added. Afterward, bound phages are eluted by different ways, including low pH, high pH, high salt concentration, enzymatic cleavage of the displayed protein or by the addition of binding competitor that compete with the phages for binding to the target molecule [108]. Then, the eluted

phages are allowed to infect *E. coli* cells to amplify the phage pool for the following selection cycle. This method is called “biopanning”, in analogy to the gold-washing procedure where gold grains are isolated from a background of unwanted materials.

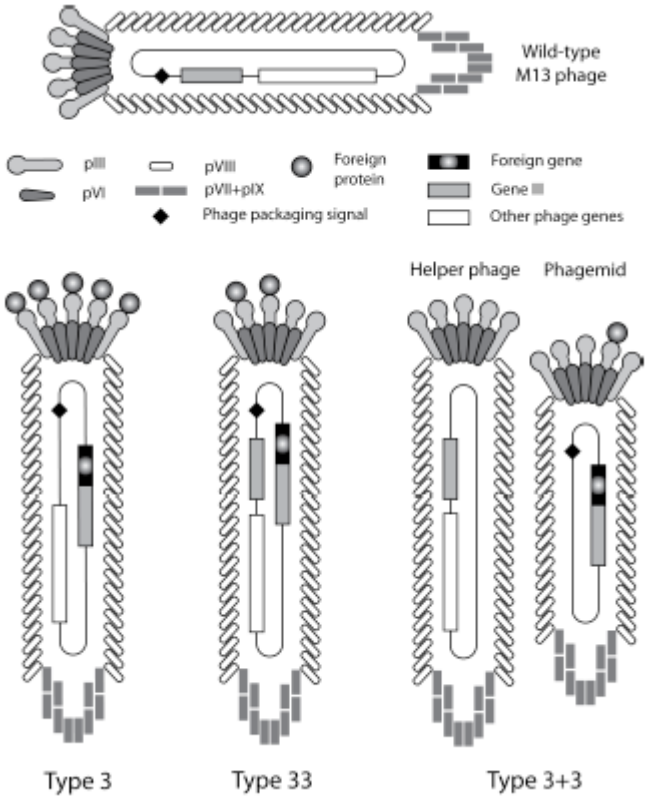


Figure 1.9. Schematic representation of M13 phage display systems. The wild-type M13 phage is encapsulated by about 2700 copies of mayor coat protein pVIII. The ends of the wild-type phage contain about 5 copies of pIII and pVI and 5 copies of pVII and pIX, respectively. The single-stranded DNA genome encodes all phage proteins and contains a phage packaging signal. Type 3 display: the phage contains one copy of the foreign gene fused to gene III and about five copies of the encoded fusion protein are displayed. Type 33 display: the phage contains both a copy of the wild-type gene III and the foreign gene fused to gene III, resulting in the display of a mixture of wild-type pIII and fusion protein. Type 3+3 display: a helper phage contains all phage genes including wild-type gene III and no functional phage packaging signal, while a phagemid contains the foreign gene

fused to gene III and a functional phage packaging signal, resulting in the display of on average no or one copy of fusion protein on the phagemid [33].

Theoretically, a single selection cycle would be sufficient to isolate binding phages from an excess of non-binding phages. However, practically, due to a presence of non-selectively binding phages, several consecutive selection cycles are typically required. After the panning procedure, individual clones are further screened and validated for their binding properties by using antigen-coated ELISA plates. The selected clones are tested in different formats, including phages, soluble proteins expressed from phagemid in periplasmic extracts or purified protein [108]. In order to avoid heterogeneity of expression levels from different clones, it would be better to sub-clone the selected clones into another expression system to prevent biased binding signals obtained from different expression level [115]. After such a preliminary screening, the promising clones with desired traits are produced, purified and further characterized for their performance such as affinity, stability and selectivity.

Because of its robustness, the possibility to generate relatively large libraries and no need of costly laboratory reagents or equipment, phage display is still the most widely used selection system.

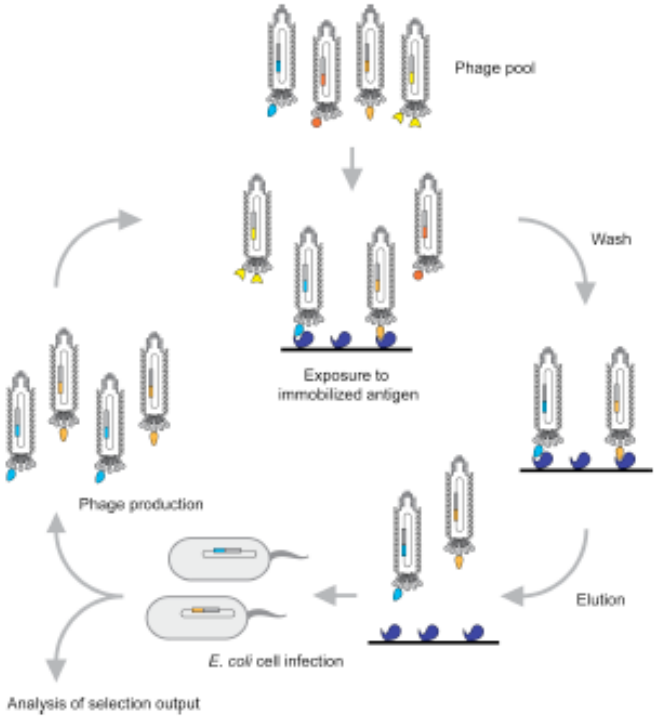


Figure 1.10. Illustration of a typical phage display selection cycle. An initial repertoire of foreign proteins displayed on phages is exposed to a target (dark purple) immobilized on a solid support. Non-binding phages are removed by washing while binding phages are eluted and allowed to infect *E. coli* cells to produce phages for an additional selection cycle (if needed) or to analyze the selected clones [33].

3.2 Cell surface display

Cell surface display involves the expression of multiple copies of a recombinant protein on a surface of microbial host, including bacteria or yeast; therefore, this offers the possibility to monitor the phenotype during selection by using flow cytometry [33].

Regarding the choice of microbial host, the most commonly hosts for cell surface display of heterologuous proteins are either yeast *(Saccharomyces cerevisiae)* or. *E. coli*. Yeast is a type of fungus encapsulated by a single membrane and a thick

layer of glucan, chitin and glycoproteins whereas *E. coli* is a gram (-) bacterium, encapsulated by two membranes and a thin murein sacculus. The functional display of a heterologuous protein on the *E. coli* outer membrane was documented in 1992 by Francisco and co-workers [116]. They fused β-lactamase to the Lpp-OmpA hybrid protein for secretion and insertion into the external surface of *E. coli*. The same research group could display a functional antibody fragment (scFv) on the external surface of *E. coli* and enrich scFv producing cells from a 10^5-fold excess of control cell background by means of FACS [117]. Thioredoxin random peptide libraries were inserted between surface-exposed sites of *E. coli* outer membrane protein LambB and flagellum protein FliC and, similarily to the biopanning procedure in phage display, selected protein-protein interactions by sequential binding and elution [118-119]. In 1988, bacterial surface display was used for a library of scFv for affinity maturation, and a clone with an approximately three-fold improved affinity could be isolated by using flow cytometry [120]. Various display scaffolds that are associated with different fusion partners have been documented, including outer membrane proteins such as OmpA, OmpX, LambB, fimbrial structure proteins such as FliC or FimH, as well as autotransporters such as AIDA-I, the IgA protease from *Neisseria gonorrhoeae* or the protease from enterohemorrhagic *E. coli* O157:H7 EspP. Contingent on the scaffold, foreign peptides or proteins can be fused to the N-terminal, to the C-terminal or inserted in loop regions.

In addition to covalent attachment, a non-covalent version for the display of full-size non-glycosylated antibodies on *E. coli* was reported by Mazor and co-workers. In their work, antibodies are secreted into the periplasm and captured on the periplasmic

face of the inner membrane by a membrane-associated Ig-binding protein. The outer membrane is permeabilized and spheroblasts decorated with antibodies can be labeled with fluorescent antigen and sorted on a flow cytometer. The spheroblasts provide only a contemporary linkage of genotype an phenotype and are not viable after FACS sorting. The selected gene pool is then transformed into fresh cells before the next sorting cycle [121].

The possibility to employ FACS for the identification of library members with improved traits may be an advantage of cell surface display over other display systems. Library size displayed on a single *E. coli* cell surface was between $5x10^4$-10^5 copies of library members. when the cell was incubated with a fluorescently labeled antigen, flow cytometry was able to be used for the detection of fluorescence signals that were directly in line with the affinity and number of the displayed library members on a cell. With multi-color FACS systems, the cells can be labeled with a second fluorescent probe with another emission spectrum that recognizes a constant part of the displayed library members. The dual-labeling allows for normalization of the fluorescence signal to the number of displayed library members, and thus fine affinity discrimination using FACS sorting. A representative aliquot of the library can be analyzed to adjust an optimal sorting gate in beforehand. Georgiou and co-workers observed that the mean fluorescence intensity resulting from different displayed scFv mutants binding to an antigen varied exactly as expected from the differences in antigen affinity in solution [122].

Hence, cell surface display can be used for the affinity determination on cell which facilitates the characterization of selected clones. Enrichment factors by using cell surface display combined with FACS sorting were reported as high as 10^6 per cycle [117]. However, there is a general limitation of cell surface display and FACS speed. A state of the art FACS can sort up to 70,000 cells per second, which practically limits the sortable library size to about 10^8 when sorting for four hours and oversampling the library ten times. Consequently, a biopanning step to enrich for productive binders and reduce the overall library size needs to be included when surpassing these numbers. Another concern is avidity effect when displaying multiple copies of library members on the same cell in combination with an oligomeric antigen. Outer membrane protein scaffolds, such as the *E. coli* LamB, may diffuse laterally along the membrane and come in close proximity, and then generating multiple binding sites to the antigen [123-124]. Another concern with *E. coli* as a host for surface display is its viability during the process of FACS sorting [125]. Comapred to yeast or gram(+) bacteria, *E. coli* does not have a thick layer of proteoglycan hence *E. coli* is susceptible to high shear forces.

As an alternative to E. coli, yeast surface could be a choice for host display. Yeast surface display was developed in 1997 by Wittrup and co-workers, who reported the selection of a scFv with improved affinity by decreasing antigen dissociation rate from a library displayed on *S. cerevisiae* [126]. The library was C-terminally fused to the Aga2p mating adhesion receptor, which enables anchoring to a-agglutinin. Per definition, the mating adhesion receptor needs to be accessible on the cell surface for haploid yeast cells to find their mating partner, which makes it very suitable for the

display of foreign proteins. Yeast possesses secretory machineries that contains chaperones, oxidative protein folding and N-linked glycosylation which makest a host applicable for the display of secreted eukaryotic proteins that require modifications such as antibody fragments, growth factors or complex cell surface receptors [127]. Since its introduction in 1997, yeast display has become popular and its potential has been demonstrated in a number of publications for affinity maturation. For example, the binding affinity of a scFv for fluorescein- biotin could be enhanced over 1000-fold to as low as 48 fM [128]. Marks and co-workers could engineer the selectivity of a botulinum neurotoxin binder to recognize both type A1 and type A2, employing a dual-selection strategy [129].

Another proposed host for display of foreign proteins is the gram(+) *Staphylococcus carnosus* which is encapsulated by a single cell membrane and a thick murein sacculus [130]. Due to thicker sacculus compared to *E. coli*, *Staphylococcus carnosus* is more resistant to high shear forces as those in the FACS nozzle; however, the efficiency of plasmid transformation is lower than that of *E. coli*. A promising approach by *Staphylococcus carnosus display* is the display of peptide libraries for the mapping of antibody epitopes which enables the identification of continuous and, to some extent, non-continuous epitopes [131-132]. Taken together, cell display systems have the main advantage in allowing for a monitoring of the phenotype to select for using flow cytometry. So far, most work was performed on yeast and *E. coli*, while *S. carnosus* may become a promising alternative.

3.3 Ribosome display

As described in the previous sections, both phage and cell display systems depend on living cells to mediate the linkage between genotype and phenotype, either directly or indirectly for the production of phages. This reliance is one major limitation of the library size because DNA has to be transformed into cells. Moreover, a "true" directed evolution experiment includes many repetitive cycles of diversification and selection. In case of cell based selection systems, the introduction of library members in between selection cycles by means of error-prone PCR or DNA shuffling involves a burdensome switch between *in vivo* selection and *in vitro* diversification, or the use of particular bacterial mutator strains such as *E. coli* mutD5 [133].

As an alternative to cell display system, a whole complete *in vitro* selection system was originally documented by Mattheakis in 1994 and further validated and improved by Plückthun and co-workers [115, 134]. The principle of ribosome display system includes a library containing DNA constructs lacking STOP-codons which are *in vitro* transcribed to generate mRNA library, which is then *in vitro* translated in a cell extract which is typically derived from *E. coli*, wheat germ or rabbit reticulocytes. Because the DNA is devoid of a STOP-codon, mRNA (genotype), ribosome and the newly synthesized polypeptide (phenotype) stay physically attached in a ternary complex. These complexes then go into a selection cycle and after each selection cycle, unbound ternary complexes are washed and removed, and mRNA molecules of captured ternary complexes are reverse transcribed and amplified in a PCR reaction to produce DNA template for the next selection cycle (Fig. 1.11).

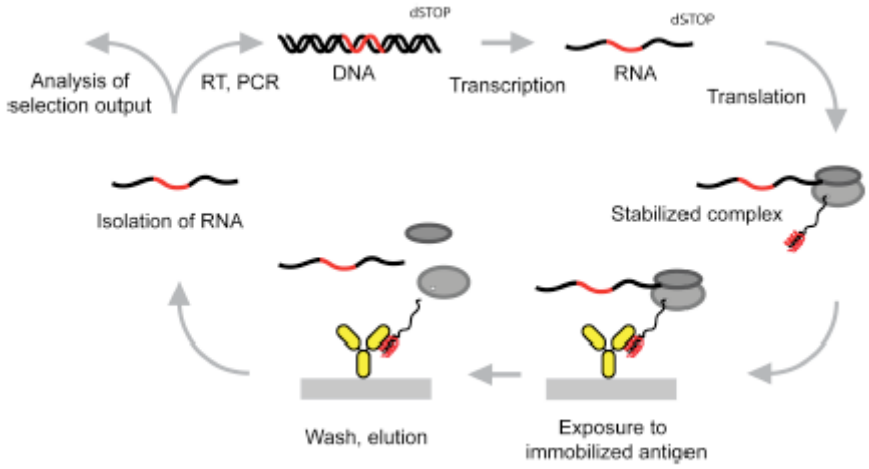


Figure 1.11. Schematic of a typical ribosome display selection cycle. A library pool of linear DNA fragments without STOP-codons is *in vitro* transcribed to yield library mRNA which is then *in vitro* translated in a cell extract. The absence of a STOP-codon and addition of Mg^{2+}-ions stalls ribosomes on their cognate mRNA and yields ternary complexes of mRNA, ribosome and displayed protein. These complexes are then exposed to an immobilized antigen; non-binding complexes are washed away and the mRNA of binding complexes is detached from complexes. Isolated mRNA is then reverse transcribed to cDNA for an additional selection cycle (if needed) or post-selection analysis. Figure is adapted from [33].

A ribosome display construct designed for use with *E. coli* platform normally consists of a T7 promoter that allows for mRNA synthesis, followed by a ribosome binding site (RBS) that can bind complementary with ribosomal RNA, thus recruiting the ribosome to the downstream start codon where protein synthesis is initiated. Contingent on the type of the cell free extract, a Shine-Dalgarno is used to intitiate translation in prokaryote while in eukaryote Kozak sequence is used to achieve translation. The open reading frame typically starts with a protein detection tag such as the RGS-His_4 tag, His_6 tag or the FLAG tag, and following by the library of binding proteins and a spacer protein. The spacer should be at least 23-30 amino acids spacer

are required for the displayed protein to exit completely from the tunnel [135]. In three publications of Ohashi, Matsuura or Osada [136-138], the SecM elongation arrest sequence was placed downstream of the spacer sequence and thereby possibly further stabilized ternary complexes. However, no comparison with classical spacer constructs has been performed so far.

The versatility of ribosome display enables the introduction of additional diversity in between selection cycles using error-prone PCR, or by DNA shuffling. This feature enables ribosome display a dynamic selection system that can mimic the natural affinity maturation process of antibodies, where somatic mutations introduce additional diversity during the maturation of B-cells. In one study, when error-prone PCR was combined with ribosome display, Hanes and coworkers reported that they could select scFv antibody fragments with equilibrium dissociation constants as low as 82 pM [115]. In another study, the combination of error-prone PCR and gene shuffling in subsequent cycles yielded an increased population of affinity-improved variants, and the highest affinity clone, as compared to error-prone PCR only [139]. Regarding phospho-specific binders, in 2014 Kummer and coworkers used ribosome display to select DARPins that specifically recognize either the unphosphory-lated (ERK2) or the phosphorylatedform (pERK2) of the kinase [28]. pERK2 binding DARPin (pE59) selected from ribosome display has binding affinity in terms of KD ~ 117 nM determined by steady-state fitting analysis with 74 of selectivity [28, 99].

3.4 mRNA display

In 1997, Roberts and Szostak invented an elegant display system called mRNA display

[140]. Similar to ribosome display, a DNA library is *in vitro* transcribed to yield library mRNA; however, in this selection platform, the mRNA library template is covalently bound to a linker of DNA or polyethylenglycol containing the translation-terminator puromycin (Fig. 1.12). During *in vitro* translation, the ribosome will reach the 3´-end of the mRNA and puromycin will enter its A-site and get covalently linked to the nascent polypeptide chain. This strategy permits a direct covalent linkage of mRNA (genotype) and protein (phenotype) in a 1:1 stoichiometry. The RNA is typically reverse transcribed to cDNA for stabilization and the complexes are then selected in a biopanning procedure. Up to date, library sizes by mRNA display as high as 10^{13} have been reported. With a huge library members, this allowed for the selection of new productive binding proteins from a completely random peptide library or naïve library or the selection of a functional enzyme from a non-catalytic library scaffold [141-142].

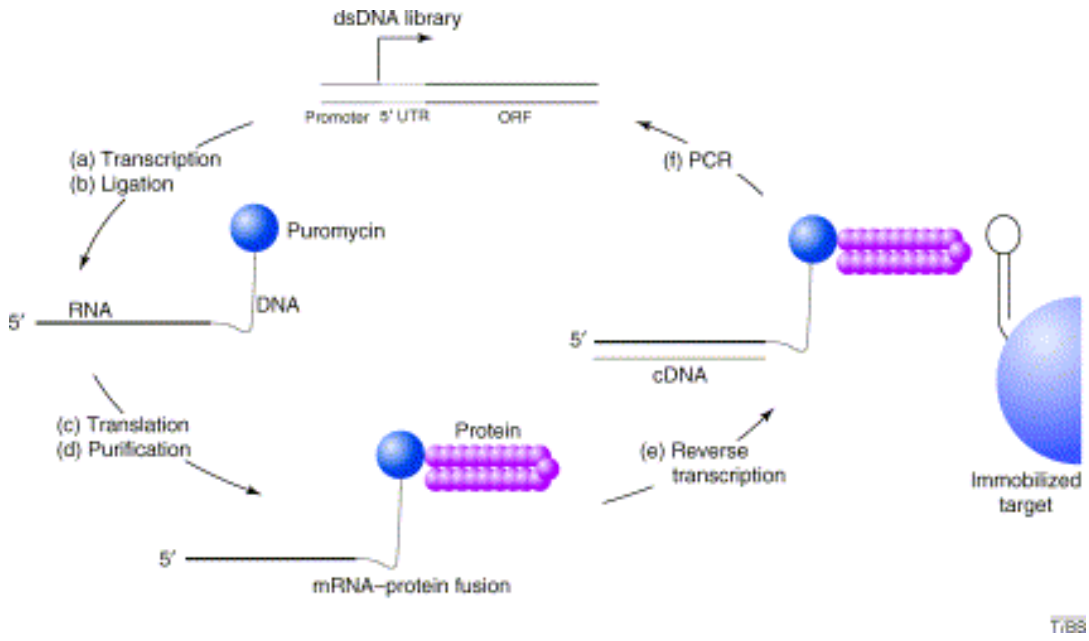


Figure 1.12. A typical mRNA-display selection cycle. (a) A library of dsDNA sequences is transcribed to generate mRNA. (b) The mRNA is ligated to a puromycin oligonucleotide (blue) and used to program an *in vitro* translation reaction. (c) cDNA synthesis is performed and (d, e) the cDNA/mRNA–protein fusion is sieved using the target of interest. (f) PCR is used to regenerate the full-length DNA construct. For targets containing RNase or RNase-H activity, the cDNA can be crosslinked to the puromycin oligonucleotide to generate a cDNA–protein fusion [142]

CIS and CAD display

CIS and CAD display are other display systems that are based on DNA-based system (Fig. 1.13). These systems have advantages of speed and stability since DNA is very stable and less susceptible to degradation compared to RNA templates [144]. CIS display involves the unique ability of a bacterial replication initiator protein, RepA, which exclusively binds back to its cognate DNA denoted cis-activity [144-145]. In principal, gene cassettes consisting of proteins of interest fused to RepA are in vitro transcribed and translated. Libraries can be quickly generated by standard PCR procedures and outputs rapidly analysed by next-generation sequencing. CIS display was the first reported to recover a specific binder from a pool of nonbinding members present at a ratio of 1 in 1010, thereby demonstrating potential for unprecedented library sizes [145]. CIS display has been also used commercially by Isogenica Ltd., UK to select high affinity peptides and folded protein domains, including antibody fragments [146]. Similar to ribosome and mRNA display, CIS display uses the benefits of cell-free systems in allowing routine access to around 10^{13} library members, while the use of a dsDNA template ensures a robust and stable system suited to simple PCR-based product recovery and library construction. Another system denoted covalent antibody display (CAD) exploits a cis-acting DNA binding protein (bacteriophage endonuclease P2A) that covalently links to its own coding strand through the activity of a catalytic tyrosine within its sequence. This system has been used for selection of tetanus toxin binders from an immune human library with enrichment rates between 14- and 300-fold [147].

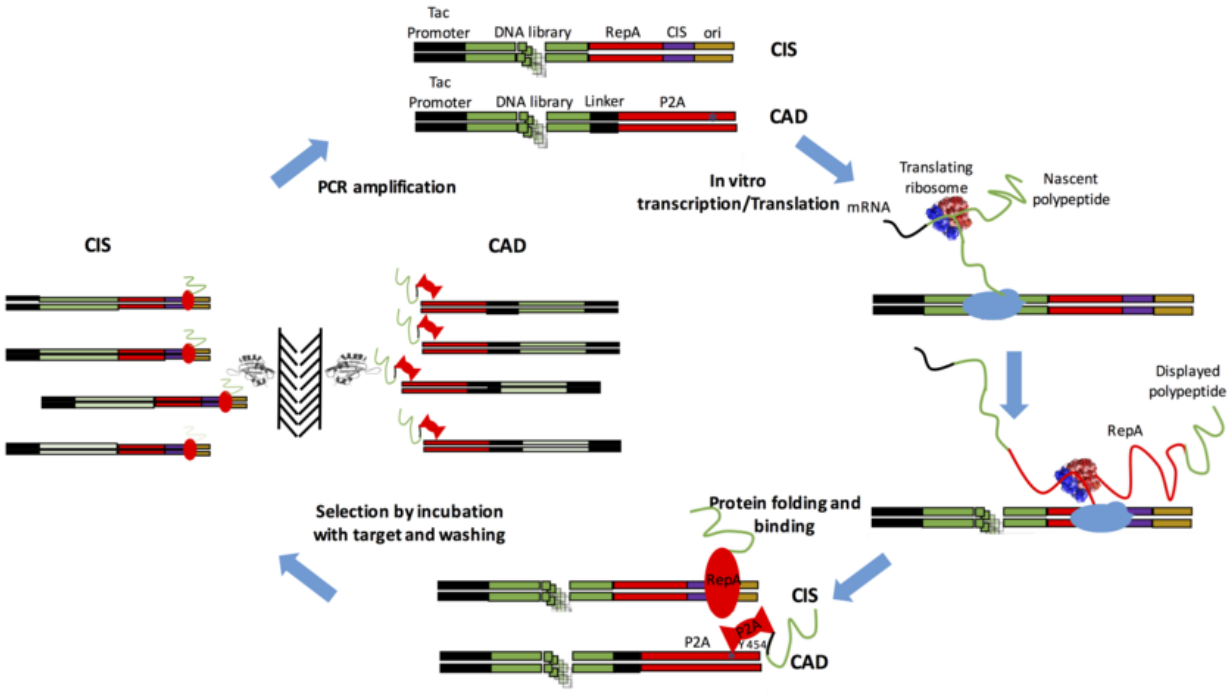


Figure 1.13. Principle of CIS display and CAD selection. In CIS display, libraries template DNA encoding an N-terminal library peptide is ligated to the RepA gene. In vitro transcription is initiated at the promoter and pauses when the RNA polymerase reaches the CIS element. Concurrent translation produces the RepA protein, which transiently interacts with the CIS element, thereby forcing its subsequent binding to the adjacent ori sequence. This process establishes a stable linkage between a template DNA and the expressed polypeptide that it encodes. In CAD (Covalent Antibody Display), mainly developed for the display of antibodies, is based on the ability of endonuclease P2A to become covalently attached (via Y454) to the 5′ phosphate of its own DNA at Ori of replication (CCT CGG, *) at position 1860. CIS display and CAD selections begin with the construction of a peptide-encoding DNA library followed by in vitro transcription/translation to form a pool of protein–DNA complexes (only shown for CIS display). The library pool is incubated with an immobilized target, nonbinding peptides are washed away, and the retained DNA that encodes the target-binding peptides is eluted and amplified by PCR, to form a DNA library ready for the next round of selection. After three to five rounds of selection, recovered DNA is cloned into an appropriate expression vector for the identification of individual target-binding peptide sequences [141].

3.5 Protein fragment complementation assay (PCA).

Another protein selection can be accomplished by utilization of natural compartments, cells, for selection of protein traits, as recognized in the protein fragment complementation assay (PCA). In PCA system, a reporter protein is divided into two components that are fused to a bait protein and a scavenger protein, respectively. Both fusion proteins are expressed within the same *E. coli* cell and once the bait binds to the scavenger, the two halves of the reporter protein are brought into contact and reconstitute its function, including fluorescence (split GFP), antibiotic resistance (split β- lactamase) or synthesis of nucleic acids and amino acids (split dihydrofolate reductase) [148]. The bait could be an antigen and the scavenger could be a library of binding proteins. Split GFP systems enables affinity selection using FACS sorting as shown in Fig. 1.14 [149], while split β-lactamase and dihydrofolate reductase allow for affinity selection using Darwinian survival of bacteria in the presence of antibiotic or absence of an essential amino acid [150-151]. In principle, the combination of PCA and Darwinian survival selection is appealing because it eliminates the requirement of antigen purification, nor repeating cycles of biopanning or FACS.

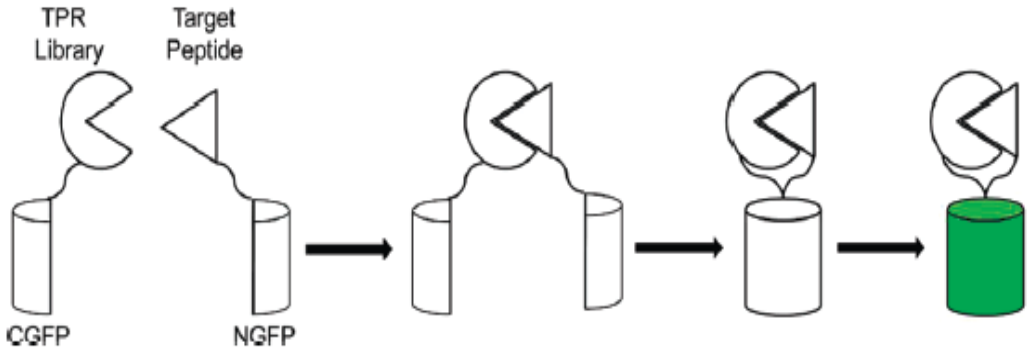


Figure 1.14. Split-GFP reassembly assay as a useful tool for screening protein-protein interactions. Green fluorescent protein is splited into two fragments. These fragments are fused to either the TPR library (CGFP) or the target sequence (NGFP). When the target peptide binds a representative TPR variant, they initiate the refolding of the GFP fragments. The two halves of GFP are then trapped and can be selected for upon fluorescence maturation. Figure is adapted from [149].

3.6 FLI-TRAP (Functional Ligand-binding Identification by Tat based Recognition of Associating Proteins) as a promising alternative to existing display technologies

As described in previous section, in vitro display selection, including phage, cell, ribosome, mRNA, and CIS display, these systems are not only helpful for identifying interaction partners for specific protein targets but they can also be used in combination with mutagenesis or randomization strategies to study the details of biologically important interactions or to engineer designer binding proteins with desired traits, including binding affinity, catalysis, or stability. However, a major drawback to all of the synthetic library approaches reported to date is that they all rely on *in vitro* selection methods which are technically demanding and labor intensive, and are implemented in cell-free environments that may not accurately reflect the complex conditions inside of a cell.

To address these shortcomings, DeLisa group has devised an ingenious and elegant technology denoted FLI-TRAP which is a bacterial-based selection method that is completely *in vivo* selection. This technology could also be superior to other display technologies because FLI-TRAP: (1) permits detection of protein interactions directly in cells, (2) may not require multiple enrichment rounds of affinity purification and amplification, (3) may be compatible with certain proteins (i.e., large proteins, cytoplasmic proteins) that are not readily displayed on phage particles, (4) requires only

selective plating of bacteria on solid medium to uncover productive binders, and (5) eliminates the need for purification or immobilization of the protein target, especially posttranslational-modified proteins, including phosphorylated proteins. FLI-TRAP is a novel genetic selection based on the hitchhiker mechanism of the bacterial Tat pathway. The power of this approach is that it directly links the *in vivo* interaction of two proteins to antibiotic resistance as a phenotypic readout. Importantly, the approach requires no prior structural or functional knowledge and allows selection of proteins with desired properties [152].

FLI-TRAP was developed based upon a hallmark of the Twin-arginine translocation (Tat) pathway: its unique ability to transport folded proteins across tightly sealed, energy-transducing membranes, notably the bacterial cytoplasmic membrane and the chloroplast thylakoid membrane [153-157]. Substrates of the Tat pathway contain N-terminal signal peptides that possess a highly conserved S/T-R-R-X-F-L-K (where X is any polar amino acid) motif [158-161]. In *E. coli*, the Tat system operates alongside the well-characterized Sec pathway and transports two broad classes of protein: globular proteins that fold too rapidly for the Sec system to handle [162] and proteins that are obliged to fold before transport such as those that bind protein subunits [163-164] or attach redox cofactors (e.g., FeS clusters, molybdopterin centers) [161] in the cytoplasm as shown in Fig 1.15. The minimal set of components required for Tat translocation consists of the integral membrane proteins TatA, TatB, and TatC [158, 165-167]. The TatA and TatB components are single-span integral membrane proteins while TatC has been shown to contain six transmembrane spans [168]. Substrate binding to TatBC appears to trigger the recruitment of TatA oligomers [169],

and forming a variable diameter ring structure that may possibly serve as a protein-conducting channel or a patch that facilitates translocation by local destabilization of the bilayer [170-171].

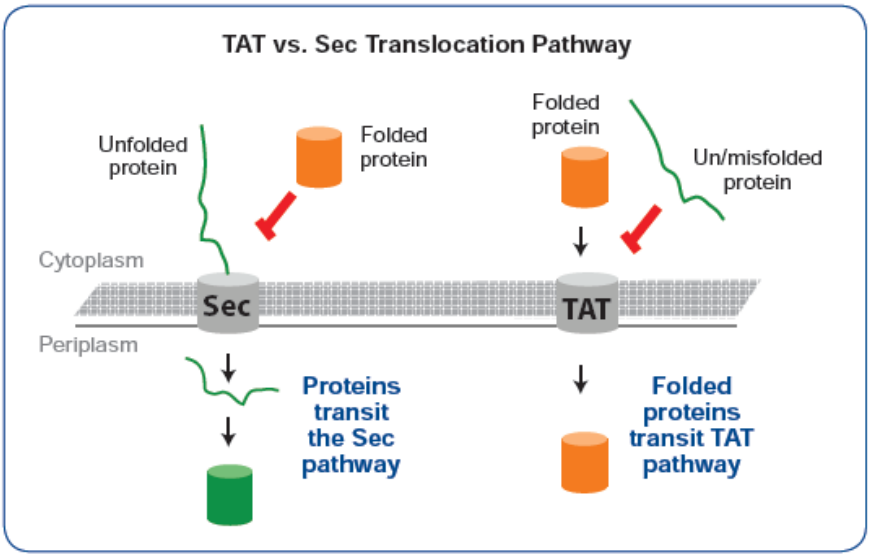


Figure 1.15. Schematic of Tat and Sec translocation pathway. In Sec pathway, protein is translocated across cell membrane to periplasm. In Tat pathway, protein need to be folded prior translocation across cell membrane from cytoplasm to periplasm.

Studies from DeLisa group have firmly established that the Tat pathway preferentially exports folded forms of globular proteins while unfolded proteins or globular proteins that expose hydrophobic patches are not compatible with transport [164,172-177]. These results led us to postulate that a folding quality control mechanism must be an inbuilt feature of the Tat translocon [172]; however, the specific details of this quality control process remain a mystery. Nonetheless, the Tat system appears to hold great promise for the expression of secreted proteins and for protein engineering applications. For instance, some recombinant proteins that fail to reach a biologically active form when targeted to the Sec pathway can achieve a functional form by targeting

to the Tat pathway, as exemplified by the green fluorescent protein (GFP) [176-177]. Moreover, because the Tat machinery does not typically accommodate unfolded proteins, purification from the periplasm can give rise to relatively pure and highly active proteins in one step as highlighted by recent work from our laboratory [174].

Along similar lines, our group recently developed a reliable genetic selection strategy for isolating interacting proteins using the Tat pathway [164]. This method is based on "hitchhiker" export of heterodimeric Tat substrates such as the endogenous *E. coli* hydrogenase-2 (HYD2) complex, comprised of subunits HybO and HybC [163]. The large HybC subunit is devoid of any known signal peptide but was reportedly exported to the periplasm via interaction with its partner, HybO, that bears an N-terminal Tat signal peptide (Fig. 1.16a). Building on this natural example, DeLisa et al. [172] demonstrated that a F_{AB} heavy-chain fused to the Tat-dependent TorA signal peptide (ssTorA) assembled in the cytoplasm with its cognate light chain lacking an export signal, and the heterodimeric F_{AB} complex was exported to the periplasm by the Tat pathway.

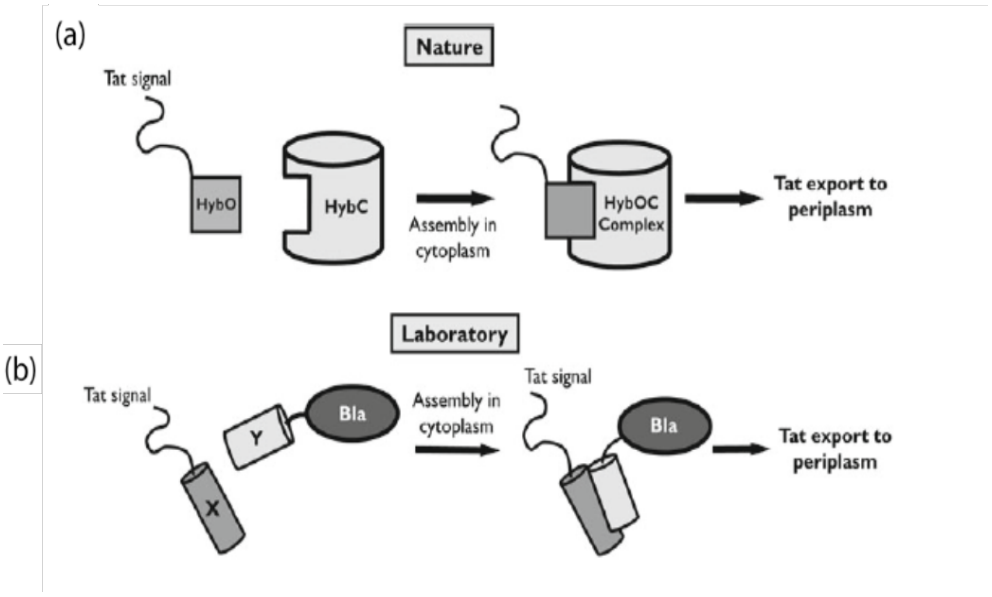


Figure 1.16. FLI-TRAP technique for genetic selection of intracellular protein-protein interaction. Schematic of the FLI-TRAP technique. (a) The method is inspired by the natural hitchhiker export mechanism of the E. coli hydrogenase-2 complex. (b) In FLI-TRAP, the hitchhiker mechanism is reconstituted by modifying a POI (X) with an N-terminal Tat-specific signal peptide such as ssTorA from the E. coli trimethylamine- N- oxide reductase TorA enzyme. The partner protein (Y) is fused to the N-terminus of mature TEM-1 Bla [152].

By simultaneously leveraging the unique abilities of the Tat pathway and the hitchhiker mechanism, the FLI-TRAP system has been developed. In FLI-TRAP, two engineered polypeptides are involved: (1) the protein of interest (POI; X in Fig. 1.16) is modified with an N-terminal signal peptide (e.g., spTorA) for targeting to TatABC; and (2) the interacting partner (Y) is fused to the N-terminus of mature TEM-1 Bla (Fig. 1.16b), which acts as a reporter for protein complex transport to the periplasm. The interaction of X and Y produces a physical linkage between spTorA and Bla in the form of a heterodimeric complex, ssTorA-X::Y-Bla. The subsequent Tat-dependent co-translocation of Bla into the periplasm renders *E. coli* cells resistant to β-lactam antibiotics. By coupling heterodimer formation in the cytoplasm with transport of Bla,

simple clonal selection can be used to identify cells manifesting interacting proteins. Bla is an ideal reporter as it allows cells to survive in the presence of β-lactam antibiotics but only if it is transported into the periplasm. Hence, in the absence of an interaction, Bla remains in the cytoplasm and does not confer β-lactam resistance to cells. In addition, Bla is a relatively small (~29 kDa), monomeric protein that can be easily expressed, is not toxic to *E. coli*, and is compatible with the Tat pathway. To date, we have demonstrated FLI-TRAP for protein selection of various single chain antibody fragments (scFvs), single chain antibodies (VHH), and synthetic binding scaffolds, in particular DARPins that this thesis is based on. In combination of error-prone mutagenesis with FLI-TRAP, scFV-GCN4 variants was selected and the selected variants showed improved stability and solubility without lowering binding affinity. Likewise, FLI-TRAP was used with a challenging target which was a-synuclein (aggregation-prone target). By combining error-prone PCR and this protein selection, a mutant VHH targeting α-synuclein protein was selected with 8 fold-improved binding affinity. Importantly, the selection procedure bypasses complications associated with purification and immobilization of binding proteins or their ligands as is required for in vitro assays.

4. Present investigation

In the next 3 chapters, the three investigations that this thesis is based upon are documented, some conclusions are drawn and future perspectives are presented. All three studies are centered on protein engineering and a novel protein selection technology denoted phospho-FLI-TRAP (hereafter PhLI-TRAP) for genetic selection of phosphorylation in living cells, and the development of affinity reagents as well as reprogramming specificity of parental binder. Of primary interest has been the investigation of PhLI-TRAP technology for selections from DARPin molecule libraries. As discussed earlier in protein selection technologies, PhLI-TRAP is completely an *in vivo* selection system, performed in *E. coli* cytoplasm, based on binding affinity and specificity between DARPin and its target without the need of purification and immobilization. The versatility of the system has two important implications. Firstly, large libraries can rapidly be generated on a nucleic acid level using recombinant DNA technology, in particular by error-prone mutagenesis to generate a starting pool. Secondly, desired traits are selected from diverse members by suitable selection condition, that is, antibiotic concentrations that allow the growth of variants, but inhibit that of parental protein.

CHAPTER 2

DEVELOPMENT OF FLI-TRAP FOR A GENETIC SELECTION FOR PHOSPHO-MODIFIED PROTEINS

1. Introduction

Recently, in vitro selection technologies, including phage display, and ribosome display, have been developed, and allow the selection of specific binding agents against phosphor-modified sites on individual targets without the need of animal immunization. These technologies have opened the door for binding scaffolds other than immunoglobulins to be engineered as specific binding proteins with the potential to overcome the drawbacks of antibodies and antibody based binding proteins. Among all non-antibody scaffolds used for molecular recognition, DARPins show the potential to be used to target phospho-modified proteins that locate intracellularly because DARPins do not contain disulfide bonds. DARPins can also be expressed in soluble form with high yields in cytoplasm of living cells. Therefore, DARPins have a potential for intracellular applications such as immunodetection and immunotherapy. Indeed, using complex DARPin libraries, target-specific binders were isolated that could reliably differentiate between two [1]states of a protein post-translationally modified by phosphorylation, and were subsequently shown to be functional in the cytoplasm of eukaryotic cells [99]. Nevertheless, a drawback to all of the synthetic library technologies documented to date, such as phage display or ribosome display, is that they

[1] The results from this chapter was submitted to Nature Communications.

are depen[2]dent on in vitro selections which are technically demanding, labor intensive in particular kinase and antigen purification, and are implemented in cell-free environments that might not accurately replicate the complex condition inside of a cell. To address these shortcomings of existing in vitro selection technology, we sought to develop a high throughput technology by extending a previously developed genetic assay termed FLI-TRAP (Functional Ligand-binding Identification by Tat based Recognition of Associating Proteins) to select phosphor-specific binders directly in living cells in a manner that greatly simplifies the process by which synthetic libraries are interrogated. FLI-TRAP is a complete in vivo selection and evolution technology based on the unique ability of the twin-arginine translocation (Tat) system to efficiently co-localize noncovalent complexes of two folded polypeptides to *Escherichia coli* periplasm.

2. Phospho-FLI-TRAP strategy (PhLI-TRAP)

Here, FLI-TRAP was functionally extended for detection of phospho-specific interaction using the extracellular signal-regulated kinase 2 (ERK2), a member of the mitogen-activated protein kinase (MAPK) family, as a model system for specific modification. ERK2 activation is mediated by the upstream MAP/ERK kinase 1 (MEK1), which phosphorylates a threonine and tyrosine within a flexible surface loop that undergoes small but significant conformational rearrangements upon modification [99]. ERK2 can also be autophosphorylated at Tyr185; however, phosphorylations at both Thr183 and Tyr185 by MEK are necessary for full activation. The Plückthun group

---

The results from this chapter was submitted to Nature Communications.

has developed DARPins specific for ERK1/2, pERK2, and ERK2/pERK2 as illustrated in Table 2.1.[3]

Table 2.1. ERK2 and pERK2 binding DARPins developed by Plückthun group

Target	DARPin	K_D, M: ERK2	K_D, M: pERK2
ERK2	E40	6.6×10^{-9}	1.2×10^{-6}
pERK2	pE59	54.5×10^{-9}	10.5×10^{-9}

To adapt FLI-TRAP for genetic selection of phosphorylation, a two-plasmid system version of FLI-TRAP (Functional Ligand binding Identification by Tat-based Recognition of Associating Proteins) was created as illustrated in the schematic shown in Fig. 2.1. The modified genetic assay called phospho-FLI-TRAP (hereafter PhLI-TRAP) is comprised of one plasmid that encodes an spTorA-pE59 chimera with N-terminal RGS-His tag for detection and a second plasmid that encodes a ERK2-reporter fusion protein, where the reporter in this context is TEM-1 β-lactamase (Bla). The two vectors are co-transformed into *E. coli* and expression of the bait and prey fusion proteins are induced. We hypothesized that the interaction of spTorA-pE59 and ERK2-Bla under $MEK1^{R4F}$ (a constitutively active mutant of human MEK1) would produce a physical linkage to form a heterodimeric complex, spTorA-pE59::pERK2-Bla and resulting in Tat-dependent cotranslocation of Bla into the periplasm in which Bla

The results from this chapter was submitted to Nature Communications.

renders *E. coli* cells resistant to β-lactam antibiotics. We can use a simple plate selection to identify cells carrying interacting proteins. In the absence of an interaction, Bla will remain localized in the cytoplasm and cells will be sensitive to β-lactam antibiotics.

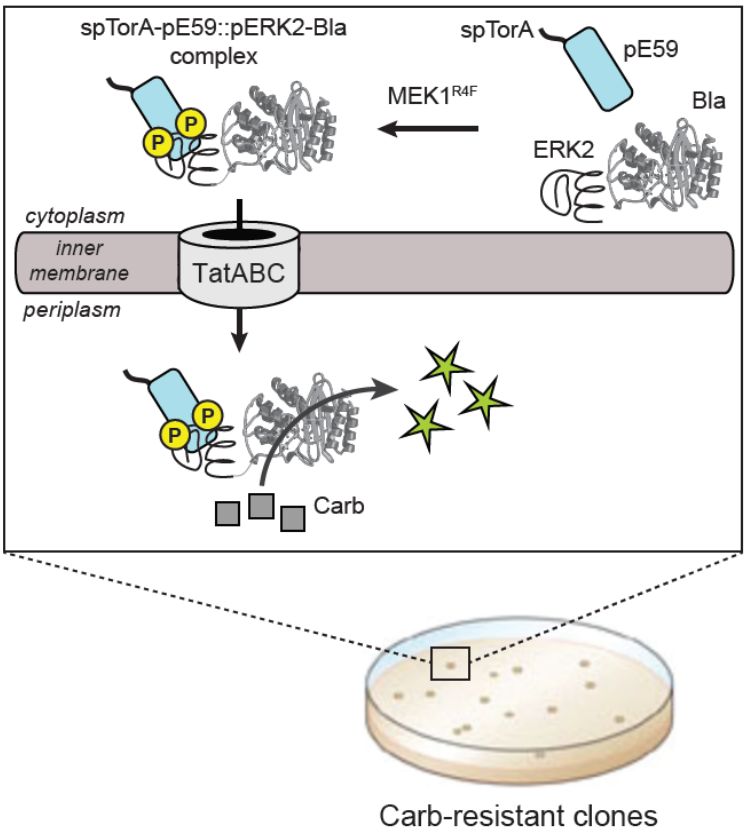


Figure 2.1. Tat-mediated hitchhiker export in bacteria. Schematic of engineered assay for co-translocation of interacting pairs via the Tat pathway is illustrated. The Tat signal peptide chosen is spTorA, the reporter is β-lactamase. pE59 is fused to spTorA while ERK2 is fused to β-lactamase. MEK1R4F is expressed for cytoplasmic phospho-modification of ERK2.

3. Validation of PhLI-TRAP strategy

To validate the PhLI-TRAP selection strategy, I employed DARPins against either the unphosphorylated or the doubly phosphorylated form of the MAPK ERK2 (ERK2 or pERK2, respectively). ERK2 is activated by phosphorylation on Thr183 and Tyr185 residues, which is catalyzed by MEK1 [34]. Specifically, DARPin pE59, which binds selectively to pERK2, was cloned into a low-copy plasmid that introduced the N-terminal Tat signal peptide derived from trimethylamine *N*-oxide reductase (spTorA) for targeting to TatABC and a N-terminal RGS-His tag for convenient detection. In parallel, a second plasmid was created in which human ERK2 was genetically fused to the N-terminus of mature TEM-1 Bla, which acts as a selectable reporter for transport to the periplasm. To generate phosphorylated ERK2 in the cytoplasm, the gene encoding a constitutively active mutant of human MEK1, namely $MEK1^{R4F}$, which is capable of activating ERK2 when expressed in *E. coli* (34), was cloned bicistronically into the low-copy DARPin expression plasmid. We hypothesized that co-expression of spTorA-pE59, ERK2-Bla, and $MEK1^{R4F}$ would result in the formation of a heterodimeric complex between spTorA-pE59 and phosphorylated ERK2-Bla (pERK2-Bla) in the cytoplasm, which would subsequently be co-translocated to the periplasm according to the "hitchhiker" mechanism as shown in Fig. 2.1 (35). Importantly, export of Bla to the periplasm renders *E. coli* cells resistant to β-lactam antibiotics, thereby enabling simple clonal selection to discriminate phospho-specific interactions.

In line with our hypothesis, co-expression of these three constructs in wild-type *E. coli* MC4100 cells resulted in $MEK1^{R4F}$-dependent phosphorylation of ERK2-Bla as depicted in Fig. 2.2 and concomitant co-translocation of the phospho-modified substrate

to the periplasmic space (Fig. 2.2) as confirmed by Western blot analysis. We observed no significant translocation of pERK2-Bla when pE59 was replaced with the well characterized DARPin OFF7, which is specific for maltose-binding protein (MBP) (28), confirming the specificity of pE59 for its cognate form of ERK2. Likewise, there was no significant translocation of pERK2-Bla in the presence of an export-defective mutant construct, spTorA(KK)-pE59, in which the essential twin-arginine residues of the N-terminal Tat signal peptide were mutated to twin lysines thereby abolishing export (Fig. 2.3).

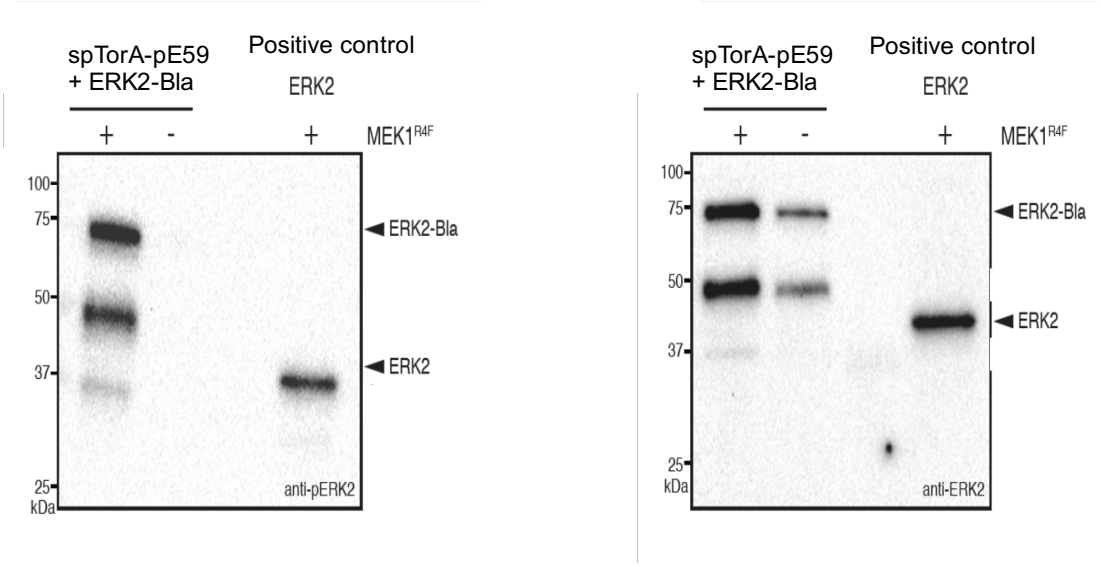


Figure 2.2. Cytoplasmic phosphorylation of ERK2 by MEK1^{R4F}. Western blot analysis of ERK2 phosphorylation when spTorA-pE59, ERK2-Bla and MEK1^{R4F} were expressed in the cytoplasm of *E. coli* MC4100.

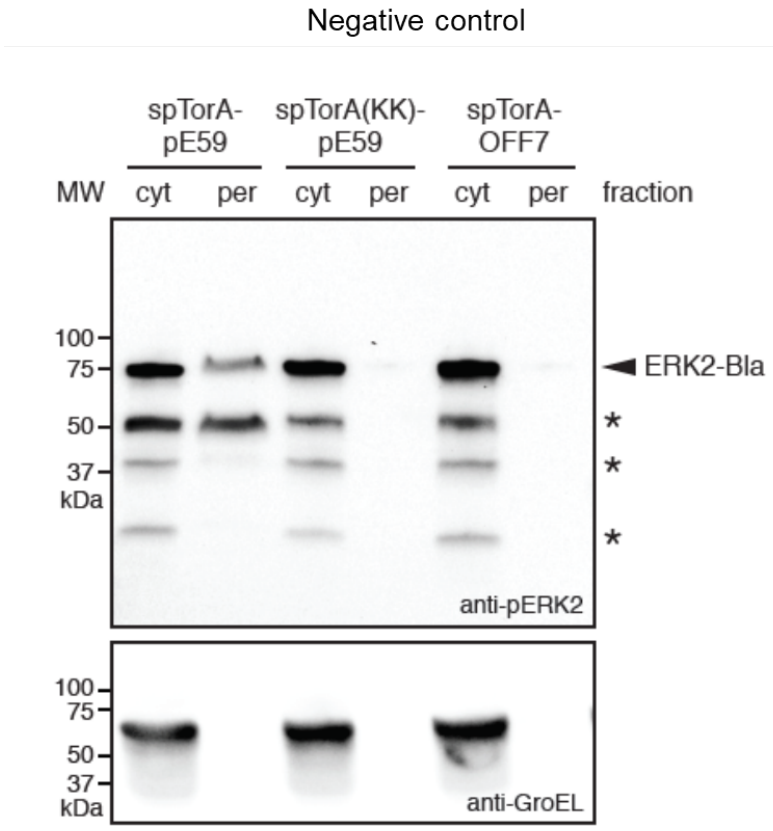


Figure 2.3. Co-translocation of pERK2-Bla chimeras to the periplasm. Western blot analysis of periplasmic (per) and cytoplasmic (cyt) fractions generated from *E. coli* MC4100 cells carrying pDD18-spTorA-pE59:MEK1^{R4F} and co-expressing pDD322-ERK2-Bla. Detection of ERK2-Bla constructs in periplasmic fraction was with anti-pERK2 antibody which detects only double phosphorylated ERK2. GroEL was detected using an anti-GroEL antibody and served as a fractionation marker.

When cells that exported pERK2-Bla to the periplasm were analyzed by spot plating analysis, we observed strong carbenicillin (Carb) resistance to a level that was even greater than that observed for positive control cells co-expressing OFF7 with MBP-Bla (Fig. 2.4). In contrast, negative control cells co-expressing a Bla fusion involving the c-Jun N-terminal kinase 2 (JNK2), a MAPK that is highly similar to ERK2, exhibited little to no Carb resistance in the presence or absence of MEK1^{R4F} (Fig. 2.4), consistent with the known specificity for pE59 (30). Importantly, the

resistance conferred by pE59, but not OFF7, was dependent on MEK1^{R4F} co-expression, indicating that the selectivity of pE59 for pERK2 over ERK2 was maintained.

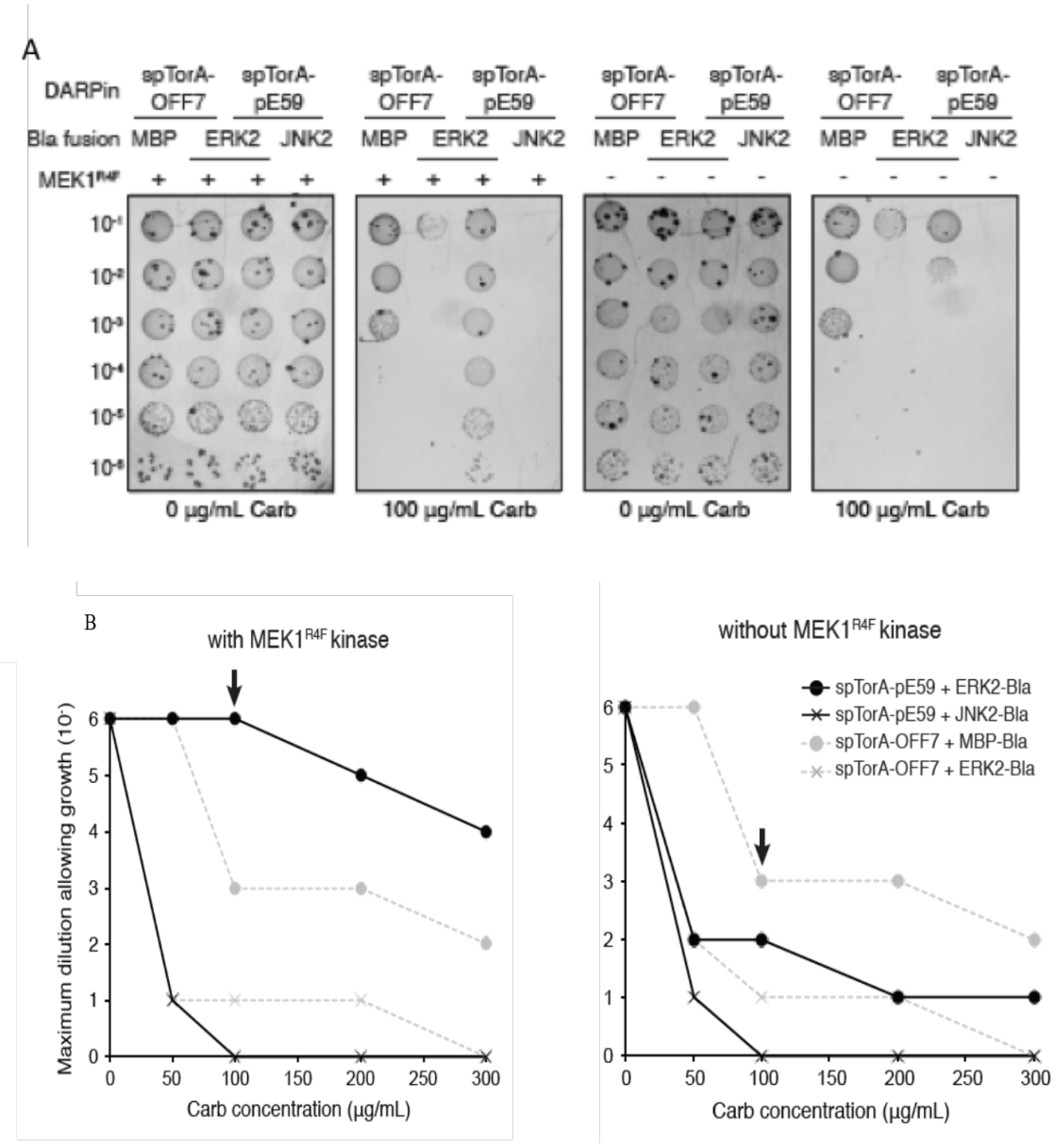


Figure 2.4. Spot dilution experiment of pE59 and OFF7 DARPin. (A) Representative spot titers for serially diluted *Escherichia coli* MC4100 cells co-expressing TatABC along with one of the following ERK2-Bla, MBP-Bla, and JNK2-Bla and one of the following DARPins: spTorA-pE59 and spTorA-OFF7 under with or without MEK1R4F

expression. Overnight cultures were normalized to OD600 of 2.5, and serially diluted in liquid LB and plated on LB agar supplemented with Carb. (B) Maximal cell dilution allowing growth is plotted versus Carb concentration (0-300 μg/ml of Carb)

Along similar lines, we found that the resistance conferred by pE59 was reduced to background when the kinase-essential phosphorylation sites at Thr183 and Tyr185 of ERK2 were mutated to Glu and Phe, respectively (Fig. 2.5), providing further support of phospho-selectivity. Collectively, these results confirm that both the high specificity and selectivity of pE59 for the phosphorylated form of ERK2 was retained in the genetic selection.

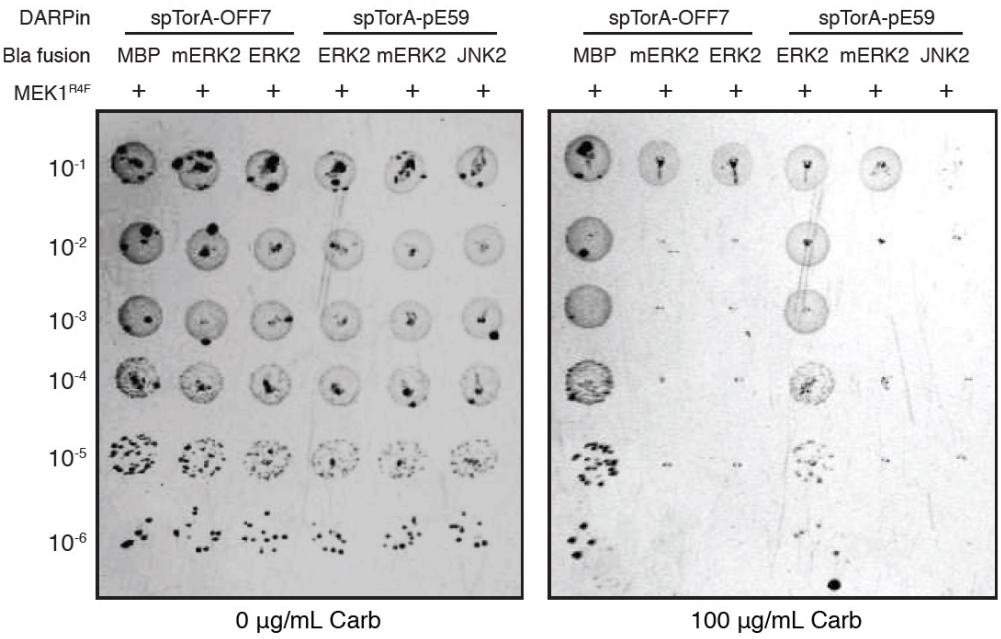


Figure 2.5. Spot dilution experiment of pE59 and OFF7 DARPin. Representative spot titers for serially diluted *Escherichia coli* MC4100 cells co-expressing TatABC along with one of the following MBP-Bla, mERK2-Bla, and ERK2-Bla and one of the following DARPins: spTorA-pE59 and spTorA-OFF7 under $MEK1^{R4F}$ expression.

Thr183 and Tyr185 of ERK2 were mutated to Glu183 and Phe185 to generate mERK2 to block double phosphorylation at both residues. Overnight cultures were normalized to OD600 of 2.5, and serially diluted in liquid LB and plated on LB agar supplemented with Carb.

4. Discussion

In this study, I have successfully developed a versatile genetic selection for protein-protein interaction called PhLI-TRAP, in particular binding against phospho-modified target (pERK2) in bacteria based on the hitchhiker [163] and folding quality control [172-174] mechanisms of the Tat transport system. By coexpression of ERK2-Bla and its kinase ($MEK1^{R4F}$) in *E. coli* cell, ERK2 could be phosphorylated in E. coli cytoplasm [13]. From spot dilution experiments, it was confirmed that Tat machinery is capable of transporting spTorA-pE59:pERK2-Bla complex when MEK1R4F was coexpressed in *E. coli* cytoplasm. The efficiency of colocalization of protein complex to the periplasm depends on folding quality/stability of the binder in *E. coli* cytoplasm, and its binding affinity, and selectivity for target. Furthermore, the PhLI-TRAP system could be used to identify protein-protein interaction in living cells, all without the need of purification. Moreover, similar to phage, cell, and mRNA display, it can be potentially utilized as a protein selection technology by combining with mutagenesis or randomization strategies to engineer binding proteins to select variants desired traits, including binding affinity, catalysis, or stability. Consequently, in the next chapter, I will explore the potential of PhLI-TRAP as a protein selection technology for post-translational-modified target, especially phospho-modified protein.

5. Materials and methods

5.1 Bacterial strains, growth and induction conditions

Wild-type (wt) *E. coli* strain MC4100 was used for all growth selection experiments. MC4100 cells were co-transformed with plasmids pDD322-TatABC-ERK2-Bla for increasing the copy number of TatABC translocases and with either pDD18-spTorA-RGS-6xHis-pE59::MEK1R4F or pDD18-spTorA-RGS-6xHis-pE59. Transformed bacteria were grown overnight at 37°C in Luria Bertani (LB) medium supplemented with 25 μg/ml chloramphenicol (Cm) and 10 μg/ml tetracycline (Tet). The next day, drug resistance of bacteria was evaluated by spot plating 5 μl of serially diluted overnight cells that had been normalized in fresh LB to OD600 = 2.5 onto LB agar plates supplemented with 1.0% arabinose and 25 μg/ml Cm as a control or varying amounts of carbenicillin (Carb; 50–300 μg/ml). Plated bacteria were incubated at 30°C for ~48 h.

5.2 Protein analysis

To prepare subcellular fractions for western blot analysis, 20–25 ml of induced culture was harvested and pelleted after 20 h incubation in 25 ^{0}C. Cells were resuspended in 1 ml subcellular fractionation buffer (30 mM Tris–HCl, 1 mM ethylenediaminetetraacetic acid (EDTA), 0.6 M sucrose) and then incubated for 10 min at room temperature. After adding 220 μl of 5 mM $MgSO_4$, cells were incubated for 10 min on ice. Cells were spun down, and the supernatant was taken as the periplasmic fraction. The pellet was resuspended in 220 μl phosphate buffered saline (PBS) and sonicated on ice. Following centrifugation at 14,000 rpm for 20 min at 4°C, the second supernatant was taken as the

cytoplasmic soluble fraction, and the pellet was the insoluble fraction. To prepare samples for cell lysate analysis, 10-15 ml of induced culture was pelleted and resuspended in 500 µl Bugbuster Mastermix. Samples were rotated at room temperature and then spun down at 14,000 rpm for 20 min at 4°C. The supernatant was taken as the soluble cytoplasmic fraction. Proteins were separated by Precise Tris-HEPES 4–20% SDS polyacrylamide gels (Thermo Scientific), and western blotting was performed according to standard protocols. Briefly, proteins were transferred onto polyvinylidene fluoride membranes, and membranes were probed with the following antibodies: mouse anti-FLAG M2-HRP (Sigma-Aldrich) to detect ERK2/pERK2-Bla fusion, rabbit anti-p44/42 MAPK (Erk1/2) (Thr202/Tyr204) antibody (Cell Signaling) to detect pERK2-Bla fusion, rabbit anti- p44/42 MAPK (Erk1/2) (137F5) (Cell Signaling) to detect ERK2-Bla fusion and rabbit anti-GroEL (Abcam) to detect housekeeping proteins in *E. coli* cells.

5.3 Plasmid construction

Plasmids used in this study are described in Table below

Plasmid	Description	Source
pDD18	P_{BAD} promoter; pBR322 *ori*, Cm^r	Laboratory stock
pDD322-TatABC	*E. coli* TatABC under native promoter ;pBR322 *ori*, Tet^r	Laboratory stock

pDD18-spTorA-RGS-6xHis-pE59::MEK1R4F	pE59 and MEK1^{R4F} with N-terminal spTorA signal and N-terminal RGS-6xHis tag in pDD18-Cm	This study
pDD18-spTorA-RGS-6xHis-pE59	pE59 with N-terminal spTorA signal and N-terminal RGS-6xHis tag in pDD18-Cm	This study
pDD18-spTorA-RGS-6xHis-Off7::MEK1R4F	Off7 and MEK1^{R4F} with N-terminal spTorA signal and N-terminal RGS-6xHis tag in pDD18-Cm	This study
pDD18-spTorA-RGS-6xHis-Off7	Off7 with N-terminal spTorA signal and N-terminal RGS-6xHis tag in pDD18-Cm	This study
pDD322-TatABC-ERK2-Bla	ERK2-Bla with N-terminal FLAG in pDD322-TatABC	This study
pDD322-TatABC-mERK2-Bla	mERK2-Bla with N-terminal FLAG in pDD322-TatABC	This study
pDD322-TatABC-JNK2-Bla	JNK2-Bla with N-terminal FLAG in pDD322-TatABC	This study

CHAPTER 3

PHOSPHO-FLI-TRAP APPLICATION: LIBRARY SELECTION OF pE59 VARIANTS WITH IMPROVED AFFINITY AND REPROGRAMMED SPECIFICITY

1. Introduction

Our previous studies reported in chapter 2 showed that the efficiency with which a TatABC-targeted binding protein escorts its cognate antigen-Bla fusion to the periplasm depends on both the expression/stability of the binding protein *in vivo* and its affinity/specificity for the antigen [152,164]. To leverage the ability of phosphor-FLI-TRAP to select pE59 variants with better performance in terms of binding affinity, isolation of pE59 variants should be possible by demanding growth on Carb concentrations that would otherwise inhibit the growth of cells expressing the parental pE59 clone. FLI-TRAP has proven especially useful for the high-throughput selection of single-chain Fv (scFv) antibodies that bind strongly to their cognate protein antigens in the intracellular environment [164]. In previous work, we improved the performance of scFv GCN4 by using error-prone mutagenesis, and FLI-TRAP for rapid isolation of intracellular antibodies (intrabodies) in the scFv format. scFv GCN4 variants possessed superior traits simply by demanding bacterial growth on high concentrations of the antibiotics. By following just a single round of survival-based enrichment using FLI-TRAP, variants of an intrabody against the yeast Gcn4p transcription factor were isolated having significantly greater intracellular stability that translated to yield enhancements of >10-fold. In addition to soluble protein target like GCN4, FLI-TRAP was extended to select scFv binder that targets an aggregation prone target, α-synuclein. Likewise, an intrabody specific for the non-amyloid component region of α-synuclein

was isolated that has ~8-fold improved antigen-binding affinity. Based on previous research, these results illustrated potentials of the FLI-TRAP method for intracellular stabilization and affinity maturation of intrabodies, all without the need for purification or immobilization of the antigen. Here, an adapted version of FLI-TRAP called phospho-FLI-TRAP (PhLI-TRAP) will be extended to improve the performance of pERK2 binder (pE59) in terms of binding affinity. Since most DARPins, including those described above, have naturally high soluble expression yields in the *E. coli* cytoplasm (36), we hypothesized that pE59 variants with enhanced affinity for cognate pERK2 antigen could be readily isolated by simply demanding cell growth on Carb concentrations that would otherwise inhibit the growth of cells expressing the parental pE59 clone.

2. Library selection of enhanced affinity of pERK2 binding DARPin

To test this hypothesis, I generated an error-prone PCR library of the entire pE59 sequence by using a GeneMorph II Kit (Stratagene), and cloned this library just after the spTorA signal peptide in a low-copy expression plasmid that also included the gene encoding $MEK1^{R4F}$. Following co-transformation of wild-type MC4100 cells with the plasmid library along with the ERK2-Bla plasmid, positive clones were selected on high concentrations of Carb (300-500 μg/mL). These concentrations were chosen because they supported outgrowth of positive hits from the library but inhibited outgrowth of individual cells expressing the parental pE59 sequence. Following just a single round of survival-based enrichment using the PhLI-TRAP assay, clones pEM1, pEM2 and pEM3 were isolated. Both pEM1 and pEM2 were isolated at 300 ug/ml of Carb at cell

dilution of 10^{-5} and 10^{-6}, respectively while pEM3 was selected at at 300 ug/ml of Carb at cell dilution of 10^{-4}. To ascertain that the greater resistance conferred by these 3 clones was due to mutations in pE59 sequence and not elsewhere in the plasmid, all isolated DARPin sequences were back-cloned into the original low-copy vector and used to transform wild-type MC4100 cells carrying the ERK2-Bla plasmid. After that, spot plating of cells co-expressing these back-cloned constructs along with ERK2-Bla and $MEK1^{R4F}$ confirmed that all three positive hits conferred significantly greater Carb resistance to cells compared to that conferred by parental pE59 as illustrated in Fig. 3.1. In addition to binding affinity, I also back cloned the pEM1, pEM2, and pEM3 sequences into the original low-copy vector but without $MEK1^{R4F}$. In the absence of $MEK1^{R4F}$, the drug resistance of pEM1, pEM2, and pEM3 was dramatically reduced to a level that was comparable to parental pE59 as shown in Fig. 3.1. In conclusion, according to spot dilution results, pEM1, pEM2, and pEM3 showed much stronger Carb resistance than parental pE59 under $MEK1^{R4F}$ expression, but when spot plating was performed in a scenario without $MEK1^{R4F}$, the resistance to Carb of all 3 hits was significantly reduced. Therefore, we hypothesized that all 3 hits would show matured binding affinity while still retaining their binding specificity.

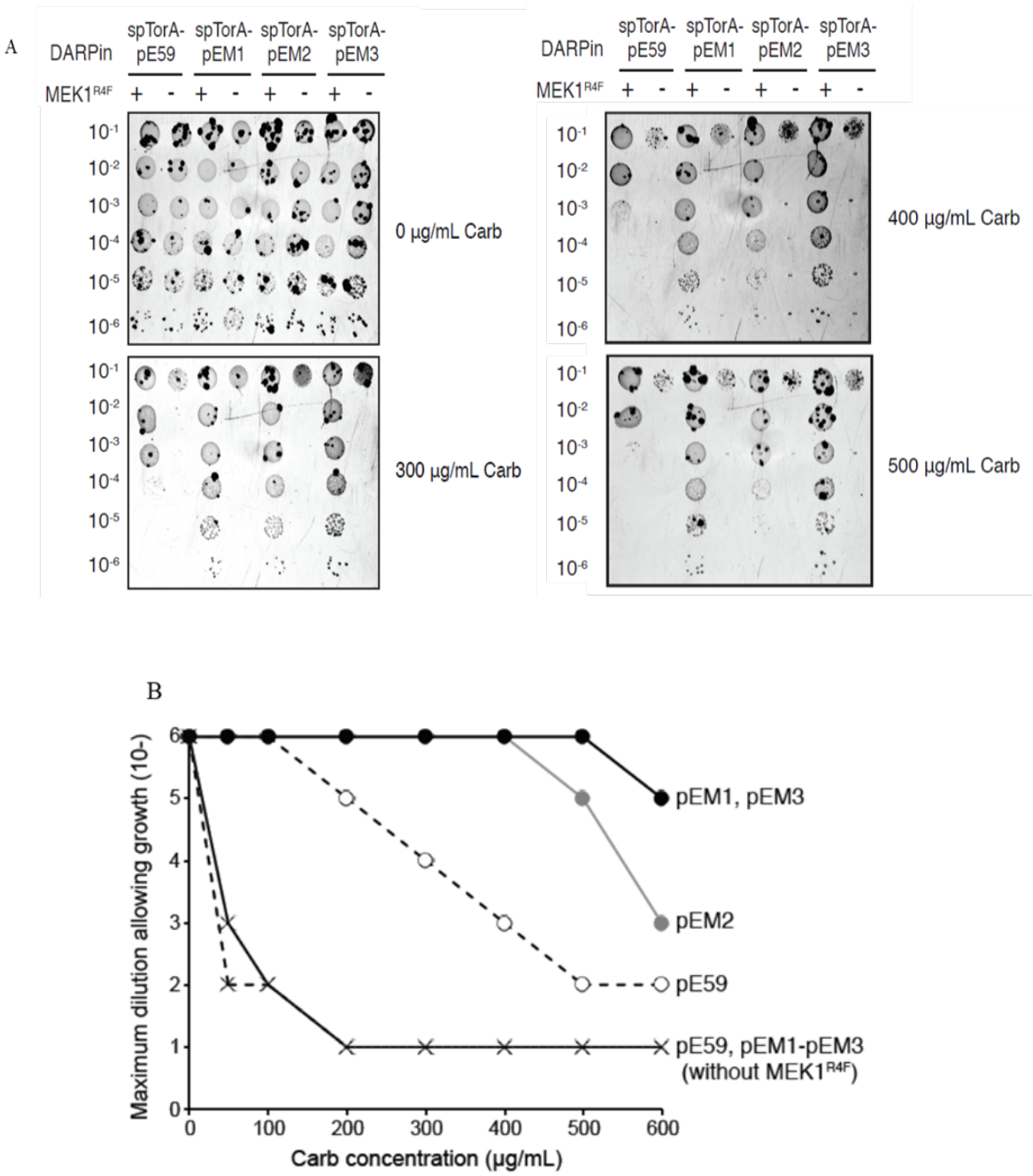


Figure 3.1. Direct selection of pE59 variants with improved affinity. (A) Representative spot titers for serially diluted *Escherichia coli* MC4100 cells co-expressing TatABC along with ERK2-Bla and one of the following: ssTorA-pE59, the parental clone, and library-selected clones pEM1–pEM3 derived from pE59 DARPin. Overnight cultures were normalized to OD600 of 2.5, and serially diluted in liquid LB and plated on LB agar supplemented with Carb. (B) Maximal cell dilution allowing growth is plotted versus Carb concentration (0-600 µg/ml of Carb)

3. Characterization of binding affinity of pEM1, pEM2, and pEM3

To prove the hypothesis of improved binding affinity of pE59 variants, pEM1, pEM2, and pEM3 sequences were subcloned in an expression vector called pDST67 with RGS-His tag at the N-terminal for purification and detection purpose. Then, I purified pEM1, pEM2, pEM3 and pE59, and the purity of all purified protein was above 95% (Fig. 3.2) which was qualified for further analysis.

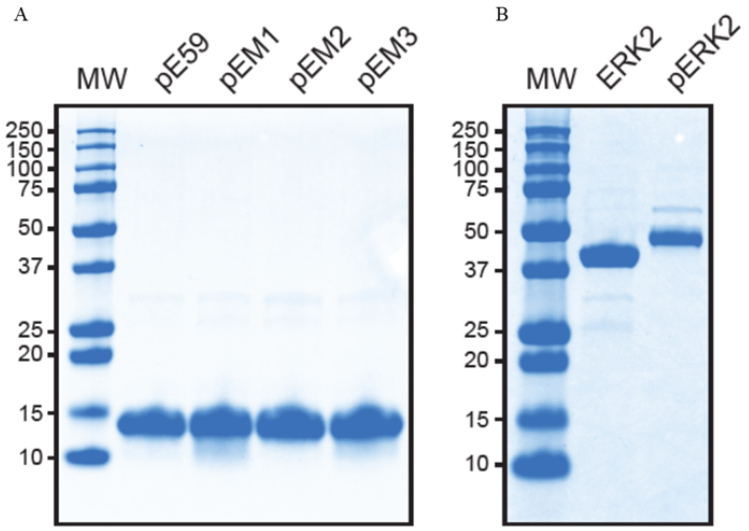


Figure 3.2. SDS-PAGE analysis and coomassie blue staining of purified proteins. (A) Coomassie blue staining of purified DARPin variants. (B) Coomassie blue staining of purified ERK2 and pERK2. The purity of all proteins in this figure was above 95% purity.

Next, I evaluated binding activity of the isolated clones by indirect enzyme-linked immunosorbent analysis (ELISA) using immobilized biotinylated ERK2 and biotinylated pERK2 as antigens. In accordance with the drug resistance results, pEM1, pEM2, and pEM3 all exhibited significantly higher binding activity against pERK2 compared to pE59 with clone pEM1 showing the greatest improvement (Fig. 3.3). When

the same clones were assayed for binding against ERK2, all showed very low binding activity that was comparable to pE59 (Fig. 3.3).

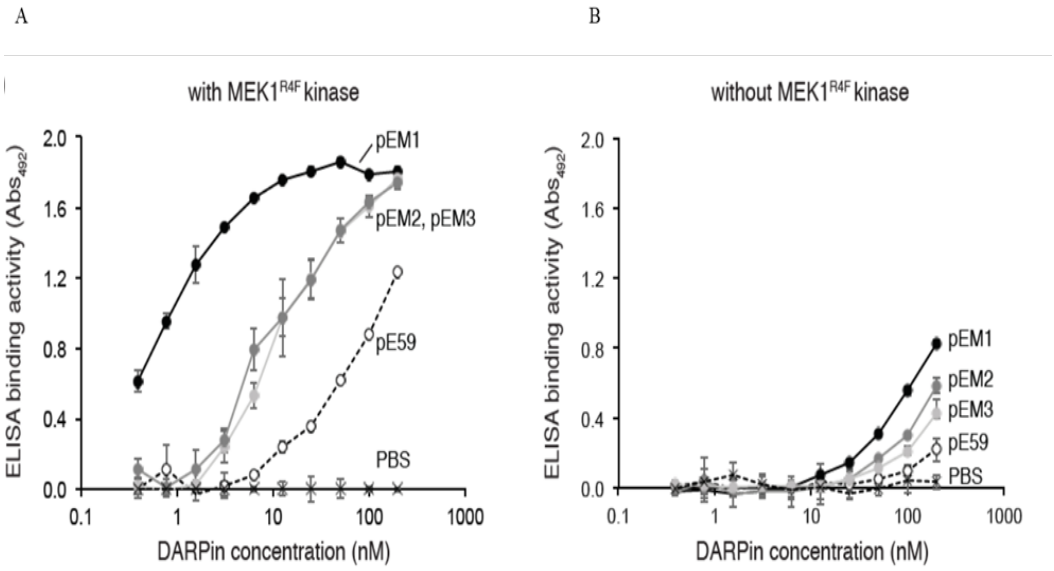
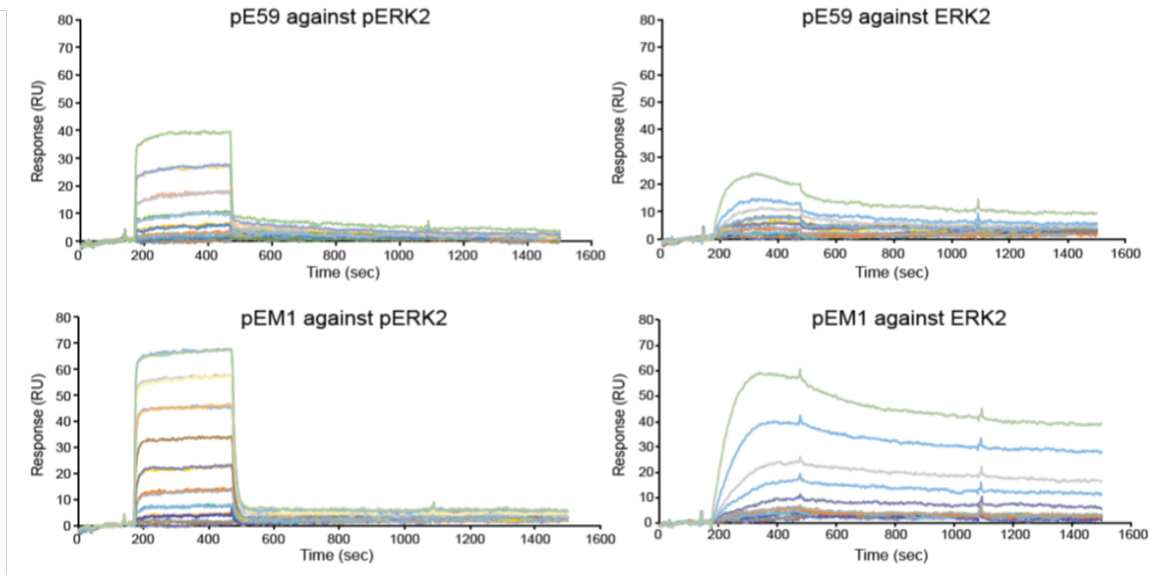


Figure 3.3. Binding specificity analysis of DARPin variants by ELISA. (A) Binding of pEM1, pEM2, pEM3, and parental pE59 against pERK2. (B) Binding of pEM1, pEM2, pEM3, and parental pE59 against pERK2. In both experiments, 100 nM of biotinylated ERK2 and pERK2 were immobilized at 4 C for an hour, and various concentrations of DARPin variants ranging from 0.39 nM to 200 nM were used for analysis of binding activity. PBS was used as a negative control for both experiments

To quantify the affinity for the most improved clone, pEM1, the equilibrium dissociation constant K_D was determined for binding to the ERK2 and pERK2 antigens by kinetic surface plasmon resonance (SPR) measurements on a Biacore instrument. To perform kinetic analysis by SPR, I immobilized the chip surface with either ERK2 or pERK2, and then flowed various concentrations of pEM1 and pE59 ranging from 0 nM to 1250 nM across the chip surface. According to the sensorgram shown in Fig. 3.4, I determined that the stronger binding measured for pEM1 in spot plating and ELISA experiments above is based on a nearly 70-fold improvement in pERK2 affinity, to 0.15

nM, while the observed selectivity in these assays is based on a >232-fold difference in K_D value for binding to cognate pERK2 versus noncognate ERK2 (Table 1, Fig. 3.4). Taken together, these results suggest that by performing genetic selections in a MEK1R4F-expressing strain background, the selectivity of the affinity-matured pE59 variants was not compromised and remained strongly biased towards phospho-modified ERK2.

Sequencing of the three hits revealed that a relatively small number of amino acid changes (F67Y in pEM1, L7M and D60G in pEM2, and L55V, N62K, and I83V in pEM3) is responsible for the increased binding affinity. Collectively, the mutations primarily mapped to the ankyrin repeat modules between the N- and C-terminal capping repeats. Interestingly, the dramatically enhanced affinity of pEM1 arises from a single mutation, F67Y, to a residue that directly contacts the surface of pERK2 (Fig. 5). According to interaction contacts in the pE59/pERK2 complex reported by Kummer et. al., Phe has interactions with 230 His, 232 Leu, and 233 Asp on pERK2; therefore, amino acid change from Phe to Tyr might generate stronger interaction between pEM1 and pERK2. For clones pEM2 and pEM3, most of the mutations (L55, N62, D60, and I83) appear to be internal to the DARPins and may not contact pERK2 directly (Fig. 5), although it is possible that these mutations alter the structure of the clustered helices in a way that enhances contact with pERK. In the case of pEM3, an additional L7M mutation in the N-capping repeat might also explain the improved binding as this residue is in position to interact with the surface of pERK2 (Fig. 3.5).



pERK2

DARPin	k_{on} (1/Ms)	k_{off} (1/s)	K_D (nM)	Chi^2
pE59	$6.1x10^4$	$6.4x10^{-4}$	10.5	0.5
pEM1	$3.2x10^5$	$4.9x10^{-5}$	0.15	3.1

ERK2

DARPin	k_{on} (1/Ms)	k_{off} (1/s)	K_D (nM)	Chi^2
pE59	$1.7x10^3$	$9.4x10^{-5}$	54.5	7.5
pEM1	$4.7x10^3$	$1.7x10^{-4}$	34.8	22.3

Figure 3.4. Binding kinetics analysis of pE59 and pEM1 against ERK2 and pERK2 using Biacore3000. ERK2 and pERK2 were immobilized at low concentrations (RU ~ 500) and the response of varied amounts of DARPins was compared with an empty flow cell. Three independent experiments were carried out for each DARPin/kinase combination. Representative results are depicted. The response of different concentrations of pE59 and pEM1 (0, 1.2, 2.4, 4.9, 9.8, 19.5, 39.1, 78.1, 156.3, 312.5, 625, and 1,250 nM) were applied to flow cells with immobilized pERK2 and ERK2 compared with an empty flow cell. Representative results are depicted. The data were evaluated by fitting the equilibrium binding responses to obtain affinity values, which are given in the table.

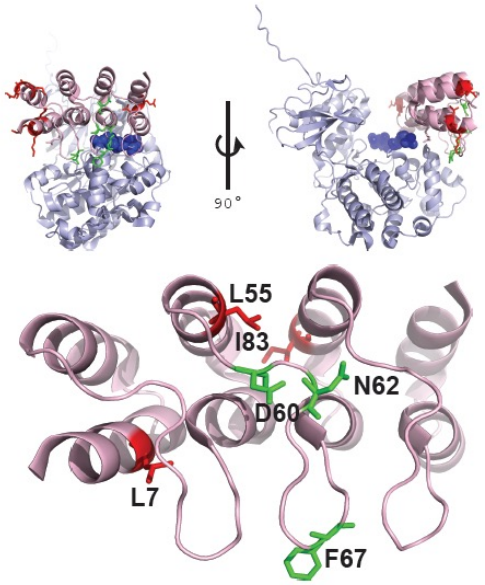


Figure 3.5. Mutations on pEM1 (F67Y), pEM2 (L7M and D60G,) and pEM3 (L55V, N62K, and I83V) mapped to the ankyrin repeat modules (pE59).

4. Library selection of reprogrammed binding specificity of pERK2 binding DARPin

According to the results documented above, I successfully isolated pE59 variants (pEM1, pEM2, and pEM3) with stronger affinity to cognate pERK2 when compared to parental pE59. Among all pE59 variants selected from PhLI-TRAP platform, pEM1 showed strongest binding affinity that could translate to 0.15 nM which was around 70-fold improved affinity. In the matter of binding specificity, pEM1 binding selectivity was about 232. Inspired and encouraged by this success of isolating pE59 variants, I next attempted to use PhLI-TRAP genetic selection to reprogram pE59 binding to nonphosphorylated ERK2. I hypothesized that pE59 sequences' library selected in the system without $MEK1^{R4F}$ expression would be able to bind better to ERK2 than parental pE59. However, due to the lack of counter selection in the selection

system, I expected to obtain pE59 hits that are promiscuous binders that are able to bind to both phosphorylated and nonphosphorylated form.

Similar to the approach outlined in the above section, an error prone library of pE59 sequences was generated and cloned just after the spTorA signal peptide in the low-copy expression plasmid; however, the gene encoding $MEK1^{R4F}$ was omitted. Following co-transformation of wild-type MC4100 cells with the plasmid library along with the ERK2-Bla plasmid, positive clones were selected on moderate Carb concentrations (50 and 300 μg/mL). At these concentrations, we anticipated that outgrowth of individual cells expressing the parental pE59 sequence would be inhibited, thereby limiting outgrowth to only positive hits from the library. Following one round of survival-based enrichment using the PhLI-TRAP assay in the absence of $MEK1^{R4F}$-mediated phosphorylation, we isolated six clones, EpEM1-EpEM6, on 50 μg/mL Carb and one clone, EpEM7 on 300 μg/mL Carb (Fig. 3.6). All isolated DARPin sequences were back-cloned into the original low-copy vector without $MEK1^{R4F}$ and used to transform wild-type MC4100 cells carrying the ERK2-Bla plasmid. Spot plating of cells co-expressing these back-cloned constructs along with ERK2-Bla, but in the absence of $MEK1^{R4F}$, confirmed that all positive hits conferred greater Carb resistance to cells compared to that conferred by parental pE59. Among all clones, two clones in particular, EpEM6 and EpEM7, stood out for their high level of Carb resistance as shown in Fig. 3.7a, which suggested that each had acquired strong binding activity toward non-cognate ERK2. Consequently, EpEM6 and EpEM7 were further investigated for their binding analysis.

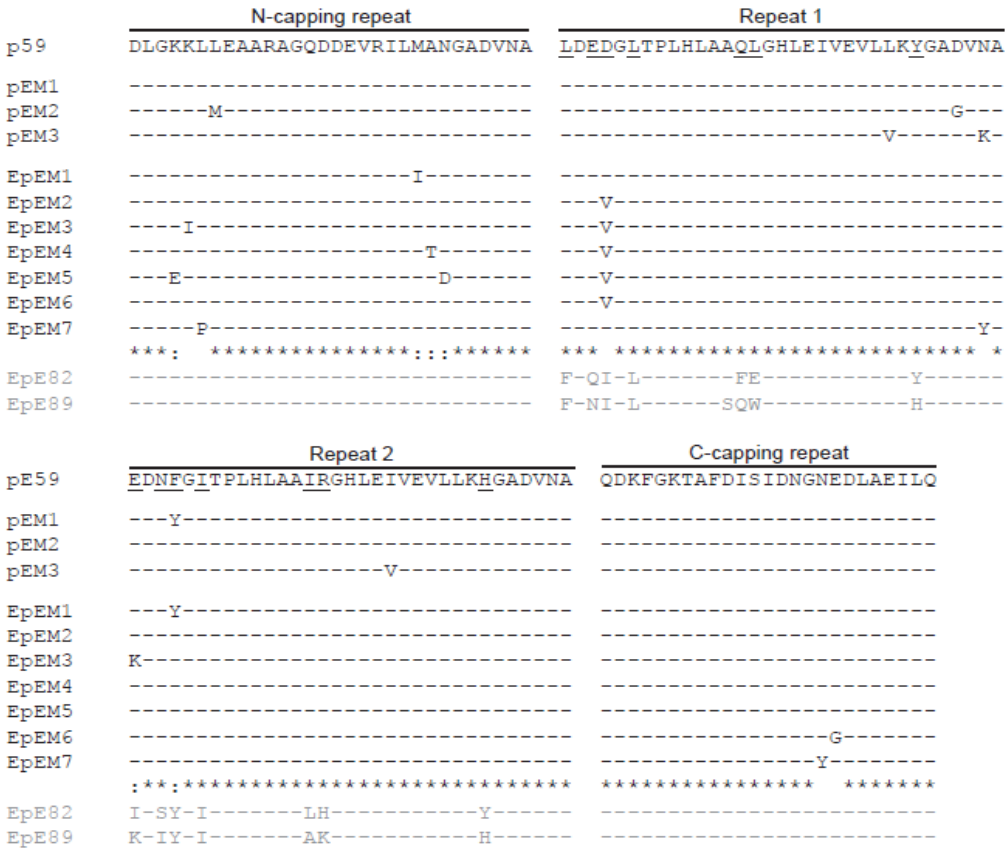


Figure 3.6. Sequence analysis of selected DARPin clones

To determine if the evolution of pE59 variants capable of binding ERK was accompanied by a relaxation of substrate selectivity, which is commonly observed in enzyme engineering studies [179], I performed spot dilution analysis of cells co-expressing EpEM6 or EpEM7 along with ERK2-Bla and MEK1^{R4F}. Clearly, both variants still conferred strong resistance to cells expressing pERK2 (Fig. 3.7a), indicating that reprogramming substrate specificity toward non-cognate ERK2 resulted in promiscuous variants that bound both ERK2 forms. Interestingly, EpEM7 and pEM1 showed a comparable level of resistance, but EpEM7 was promiscuous in binding

selectivity. Considering the mutation of DARPin variants, in particular EpEM6 and EpEM7, it is interesting to note that two of the three mutations in EpEM7, namely L6P and N62Y, are similar to the L7M, D60G, and N62K mutations uncovered in affinity-matured pEM2 and pEM3 (Fig 2b), which might explain the strengthened binding to pERK2 by EpEM7. SPR experiments were performed to quantify the affinity for for EpEM6 and EpEM7. Similar to KD determination previously analyzed for pEM1, the equilibrium dissociation constant K_D was determined for binding to the ERK2 and pERK2 antigens by kinetic surface plasmon resonance (SPR) measurements on a Biacore instrument. Either ERK2 or pERK2 was immobilized on the surface of SA chip, and then various concentrations of EpEM6 and EpEM7 ranging from 0 nM to 1250 nM were flowed across the chip surface. According to the sensorgram shown in Fig. 3.8, EpEM6 showed similar K_D for both ERK2 and pERK2 which was 10.8 nM and 10.2 nM, respectively. These results were totally in agreement with spot plating experiment of EpEM6 shown in Fig. 3.7a depicting that the resistance of EpEM6 under either with or without MEK1R4F showed similar resistance against antibiotic. Compared to EpEM6 and pE59, EpEM7 showed stronger binding affinity against ERK2, which translated to K_D of 6.37 nM while KD of EpEM7 against pERK2 (0.528 nM) was comparable to that of pEM1. Taken together, these results suggest that by performing genetic selections without $MEK1^{R4F}$, this enabled selection of pE59 variants which showed promiscuity in binding against ERK2 and pERK2.

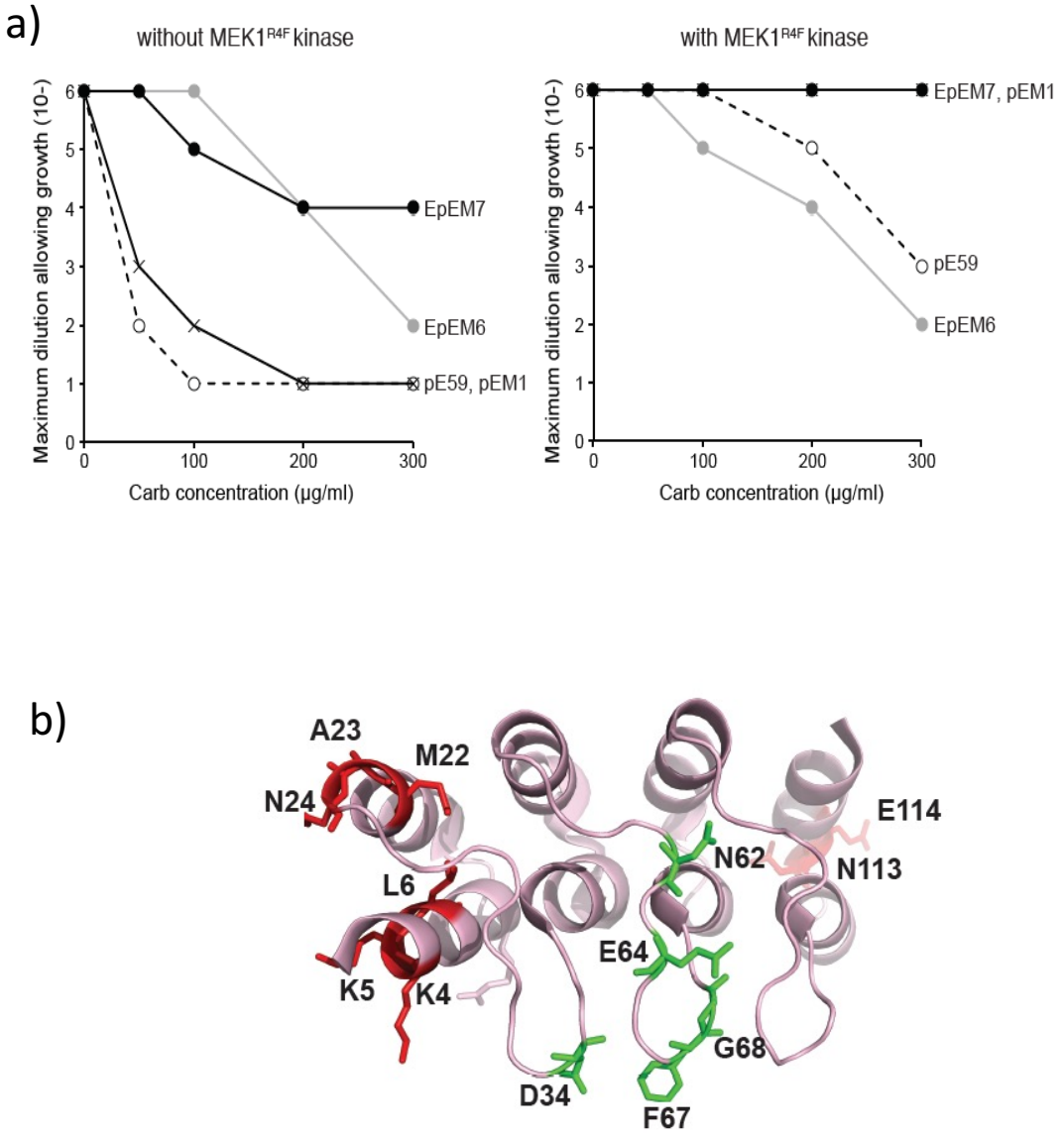
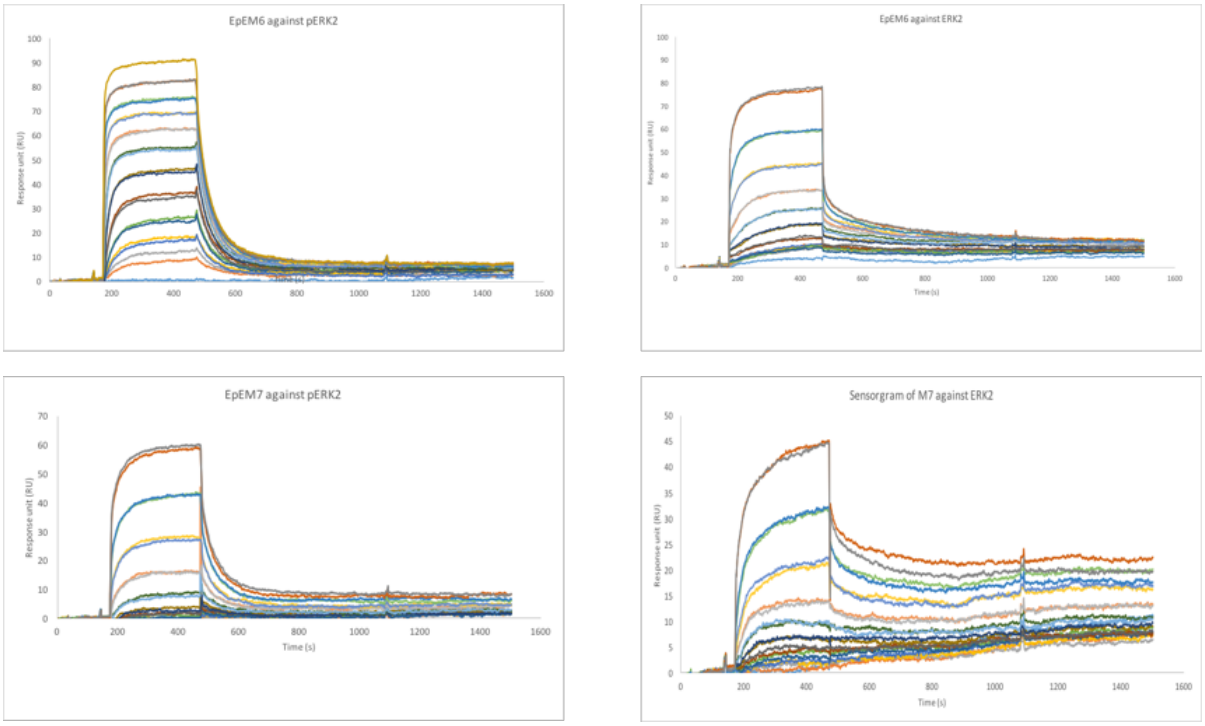


Figure 3.7. Direct selection of pE59 variants with reprogrammed binding affinity and specificity. (A) Representative spot titers for serially diluted *Escherichia coli* MC4100 cells co-expressing TatABC along with ERK2-Bla in the presence and absence of MEK1^{R4F} and one of the following: spTorA-pE59, the parental clone, and library-selected clones pEM1, EpEM6, and EpEM7 which derived from pE59 DARPin. Overnight cultures were normalized to OD600 of 2.5, and serially diluted in liquid LB and plated on LB agar supplemented with Carb. (B) Mutations on EpEM1-EpEM7 mapped to the ankyrin repeat modules (pE59).



pERK2

DARPin	k_{on} (1/Ms)	k_{off} (1/s)	K_D (nM)	Chi^2
EpEM6	$5.23x10^5$	$5.31x10^{-3}$	10.2	9.8
EpEM7	$2.3x10^4$	$1.22x10^{-5}$	0.528	8.6

ERK2

DARPin	k_{on} (1/Ms)	k_{off} (1/s)	K_D (nM)	Chi^2
EpEM6	$1.14x10^5$	$1.24x10^{-3}$	10.8	14
EpEM7	$1.39x10^4$	$8.84x10^{-5}$	6.37	5.34

Figure 3.8. Binding kinetics analysis of EpEM6 and EpEM7 against ERK2 and pERK2 using Biacore3000. ERK2 and pERK2 were immobilized at low concentrations (RU ~ 500) and the response of varied amounts of DARPins was compared with an empty flow cell. Three independent experiments were carried out for each DARPin/kinase combination. Representative results are depicted. The response of different concentrations of EpEM6 and EpEM7 (0, 1.2, 2.4, 4.9, 9.8, 19.5, 39.1, 78.1, 156.3, 312.5, 625, and 1,250 nM) were applied to flow cells with immobilized pERK2 and ERK2 compared with an empty flow cell. Representative results are depicted. The data were evaluated by fitting the equilibrium binding responses to obtain affinity values, which are given in the table.

5. Discussion

In this chapter, we first attempted to adapt PhLI-TRAP method for direct selection of DARPin variants in living cells. Based on the ability of PhLI-TRAP to differentiate the selective binding between pE59 and pERK2, but not their counterpart, I hypothesized that increasing the antibiotic selection pressure could be used to select DARPin variants with improved binding affinity because most DARPins including the one described above have naturally high soluble expression yields in the *E. coli* cytoplasm [178]. To test this hypothesis, cells expressing combinatorial DARPin libraries in the PhLI-TRAP system were challenged to grow on high concentrations of antibiotic, i.e. above the minimum inhibitory concentration, MIC, for the interaction of the parental antibody with its antigen. All isolated clones (pEM1, pEM2, and pEM3), selected from high concentration and high dilution that inhibited the growth of parental clone, exhibited intracellular stability and greater antigen-binding affinity than progenitor clones. Most importantly, all clones still retain selective binding to phosphorylated-ERK2. In all cases reported here, enhancement of DARPin traits resulted from just a handful of mutations that were acquired in just one round of laboratory evolution. It should be pointed out that additional cycles of mutagenesis and selection could be performed to further enhance stability and affinity as desired, although whether there is a ceiling that limits the affinities or stabilities achievable is an open question.

A unique aspect of this method is the natural incorporation of the Tat folding QC mechanism, which precludes export of aggregated proteins and protein complexes [171-174]. In the FLI-TRAP context, folding QC effectively eliminates poorly folded

scFv clones and allows screening for intracellular stability and antigen binding in a single step. It is noteworthy that previous attempts to use the Tat QC mechanism for optimizing intrabody folding and stability have been reported (Fisher and DeLisa, 2009; Fisher et al., 2011). In these earlier reports, an antigen-independent selection strategy was used to evolve anti-b-galactosidase scFv antibodies with greatly increased intracellular stability [181]. The power of the FLI-TRAP selection is that intracellular stability and binding affinity are assayed simultaneously in a single experiment, thereby ensuring both properties are, at a minimum, maintained in all isolated binding agents. When combining FLI-TRAP with a reconstituted MAP kinase phosphorylation cascade that promotes cytoplasmic phospho-modification of ERK2 (34), the modified genetic assay denoted PhLI-TRAP reliably reported the specificity and selectivity of an existing panel of DARPins [99] that selectively bind the nonphosphorylated (inactive) form of ERK2 or its doubly phosphorylated (active) form, pERK2. Following validation, the selection strategy was successfully used to enhance the affinity of a phospho-specific DARPin for its cognate pERK2 antigen but not their counterpart. After obtaining DARPin variants with improved binding affinity, I extended this PhLI-TRAP platform to select DARPin variants with altered specificity, which was simply performed by selecting the DARPin library pool in the absence of $MEK1^{R4F}$. Without functional expression of kinase, ERK2 phosphorylation is inhibited, so this would apply negative pressure to cells during the course of evolution to select variants with altered specificity. As expected, two DARPin variants (EpEM6 and EpEM7) show stronger binding to ERK2 compared to their parental DARPin pE59, and still retaining binding to pERK2. DARPin variants engineered by using this screening techniques to accept novel target

typically show a higher degree of binding promiscuity, which is commonly found in enzyme engineering [180].

By linking antibiotic resistance with phospho-epitope binding in the cytoplasm of *E. coli* cells, the PhLI-TRAP method eliminates the need for purification or immobilization of the phosphoprotein target and only requires selective plating of bacteria on solid medium to discover productive binding agents. PhLI-TRAP represents a simpler alternative to existing methods (animal immunization, in vitro selection technologies), offering savings in time and resources, while also providing a reliable tool for generating phospho-specific detection reagents that are both high-quality and renewable.

Moreover, beyond the identification of PTM-directed binding proteins and their subsequent engineering, we also envision other ways of applying FLI-TRAP in the future. These opportunities arise from the linkage between bacterial cell resistance and three system components: the binding protein, the post-translational modifying enzyme(s), and the substrate protein. For example, one could imagine using FLI-TRAP for high-throughput selection of synthetic libraries encoding one of these components to reveal sequence determinants that govern the activity of the post-translational modifying or that define the modified sites on a target protein. It is also conceivable that the genetic selection strategy could be reconfigured for other types of PTMs, in particular those that have been functionally reconstituted in the cytoplasm of living *E. coli* cells such as *N*-acetylation, glycosylation, neddylation, sumoylation, methylation, and ubiquitination.

6. Materials and methods

6.1 Bacterial strains, growth and induction conditions

Wild-type (wt) *E. coli* strain MC4100 was used for all growth selection experiments. MC4100 cells were co-transformed with plasmids pDD322-TatABC-ERK2-Bla for increasing the copy number of TatABC translocases and with either pDD18-spTorA-RGS-His-pE59::MEK1R4F or pDD18-spTorA-RGS-His-pE59. Transformed bacteria were grown overnight at 37°C in Luria Bertani (LB) medium supplemented with 25 μg/ml chloramphenicol (Cm) and 10 μg/ml tetracycline (Tet). The next day, drug resistance of bacteria was evaluated by spot plating 5 μl of serially diluted overnight cells that had been normalized in fresh LB to OD600 = 2.5 onto LB agar plates supplemented with 1.0% arabinose and 25 μg/ml Cm as a control or varying amounts of carbenicillin (Carb; 0–500 μg/ml). Plated bacteria were incubated at 30°C for ~48 h. Wild-type (wt) *E. coli* strain XL1-Blue was used for cytoplasmic expression of DARPins from pDST67 plasmids. Cultures were grown in LB medium supplemented with 50 μg/ml Ampicillin (Amp), and protein expression was induced with isopropyl β-d-1-thiogalactopyranoside (0.1 mM). Strain BL21(DE3) was used for cytoplasmic expression of ERK2 and pERK2 from pLK1-ERK2 and pLK1-ERK2:MEK1R4F plasmids.

6.2 Library construction and selection

A random mutagenesis library was generated from pE59 using the Genemorph II random mutagenesis kit (Stratagene). PCR was performed using 1 ng pDD18-spTorA-RGS-His-pE59::MEK1^{R4F} as template in each reaction. The resulting PCR

products were digested by XbaI and SalI, purified by gel electrophoresis, and cloned into pDD18-spTorA-RGS-His-pE59::MEK1^{R4F} that had been digested with the same enzymes. The library was transformed into electrocompetent DH5α cells and selected on LB agar containing Cm to recover clones containing the plasmid. The library size and error rate were determined to be 2×10^6 members and ~3 mutations per gene, respectively. The plasmid library was miniprepped from DH5α and used to transform electrocompetent MC4100 cells already harboring the pDD322-TatABC-ERK2-Bla plasmid. Transformed cells were incubated at 37°C for 1 h without any antibiotics and then were subcultured into fresh LB containing 25 μg/ml Cm and 10 μg/ml Tet to ensure that cells contained both plasmids. After ~16 h, cells were spun down and normalized in fresh LB to $OD_{600} = 2.5$ followed by direct plating of 100 μl of diluted cells (to the dilution factor previously determined by spot plating) onto LB agar supplemented with 1% arabinose and 300-500 μg/ml Carb. Hits were randomly picked after incubation at 30°C for ~48-72 h. An identical selection of cells carrying the pDD18-spTorA-RGS-6xHis-pE59 or spTorA-RGS-6xHis-pE59: MEK1^{R4F} (either with or without co-expression of MEK1^{R4F}) was performed as negative control. Randomly chosen positive clones were screened by spot plating to confirm Carb resistance and then sequenced to determine the identity of any mutation(s). After sequencing, the genes encoding the DARPin hits were PCR amplified, back-cloned into pDD18 with and without MEK1R4F, and used for spot plating analysis to confirm binding affinity against pERK2 and ERK2.

6.3 Library plating selection condition

Screening conditions were identified by spot plating 5 µL of overnight cells that had been normalized in fresh LB to OD600 = 2.5 onto LB agar plates supplemented with 1.0% arabinose and 0-300 µg/ml Carb and incubating at 30°C for ~ 48 hours. At Carb concentration of 200 µg/ml, pE59 could grew well at 10^{-5} dilution of overnight culture but when the antibiotic concentration was increased to 300 µg/ml, pE59 resistance was reduced to 10^{-3} dilution of overnight culture. Consequently, this concentration (300 ug/ml) and 10-4 dilution and higher concentrations were used for selection of potential hits with improved performance.

6.4 Protein purification

For DARPin purification, bacterial cells were harvested by centrifugation and the resulting cell pellets were resuspended in binding buffer (20 mM sodium phosphate, 500 mM NaCl, 20 mM imidazole, pH 7.4). The cell suspensions were then passed five times through an EmulsiFlex™-C5 cell homogenizer (Avestin; 15,000 psi/4°C) and centrifuged at 12,000 rpm for 30 min at 4°C. The clarified lysate was filtered through a 0.2-µm-syringe filter prior to sample loading. The sample was initially loaded through a 1-ml Ni-resin (GE Healthcare). The column was then washed with buffer containing 20 mM sodium phosphate, 500 mM NaCl, 60 mM imidazole, pH 7.4. The captured protein was eluted with buffer containing 20 mM sodium phosphate, 500 mM NaCl, 250 mM imidazole, pH 7.4. Final purity of proteins was confirmed by SDS-polyacrylamide gel electrophoresis (PAGE) and Coomassie staining. Purity of all proteins was typically >95%.

For biotinylated ERK2 and pERK2 purification, bacterial cell pellets were harvested by centrifugation, pelleted, and resuspended in PBS (pH 7.4) with 1 mM DTT and 0.05% Tween-20. The cell suspensions were then homogenized as above. The soluble lysate containing biotinylated ERK2 and pERK2 was first purified using avidin agarose (Thermo Scientific). The lysates were then loaded onto the packed-avidin agarose column by gravity flow. The column was washed twice with PBS buffer, after which purified fusion protein was eluted using PBS buffer containing 2 mM biotin. The eluents were passed over a Ni-column to further enhance their purity and to remove unconjugated biotin, and the proteins were eluted with buffer containing 20 mM sodium phosphate, 500 mM NaCl, 250 mM imidazole, pH 7.4. Biotinylated ERK2 and pERK2 were analyzed by SDS-PAGE followed by Coomassie staining to confirm purity, which was typically >95% for both proteins.

6.5 ELISA

Biotinylated ERK2 and pERK2 (100 nM) were immobilized on neutravidin-coated ELISA plates for 2 h at 4°C, and then washed twice with PBS (pH 7.4) with 1 mM DTT and 0.05% Tween-20. Next, the plates were blocked with PBS (pH 7.4) with 1 mM DTT, 0.05% Tween-20, and 1% (w/v) BSA. All subsequent ELISA steps were performed at 4°C in PBS (pH 7.4) with 1 mM DTT and 0.05% Tween-20. To measure binding activity, different concentrations of purified DARPins (pEM1, pEM2, pEM3, and pE59) ranging from 0-200 nM were applied to wells with or without ERK2 or pERK2 for 1 h at 4°C. DARPin binding was detected by mouse anti-RGS-4xHis antibody (Qiagen) followed by a secondary antibody/alkaline phosphatase conjugate

(Thermo Scientific), both in PBS (pH 7.4) with 0.05% Tween-20. After 1 h of incubation at room temperature, plates were washed and then incubated with SigmaFast OPD HRP substrate (Sigma) for 30 min in the dark. The reaction was quenched with 3 M H_2SO_4, and the absorbance of the wells was measured at 492 nm.

6.6 SPR

Kinetic SPR measurements were made using a Biacore 3000 instrument (GE Healthcare) for DARPins pE59, pEM1, EpEM6, and EpEM7. The running buffer was 50 mM Tris (pH 7.4), 150 mM NaCl, 0.05 mM EDTA, and 0.005% Tween-20 for pE59 and Tris (pH 7.4), 1 mM DTT, 0.05 mM EDTA, and 0.005% Tween-20 for E40 measurements. Biotinylated ERK2 or pERK2 was immobilized on a streptavidin SA chip (GE Healthcare) to ~500 response units (RU). Interactions were determined by injecting varying concentrations of each DARPin (1.2-5000 nM) at a flow rate of 30 μL/min for 5 min, after which off-rate measurements were made by flowing running buffer for 50 min. The signal of an uncoated reference cell was subtracted from the sensorgrams. Zero-concentration samples (Tris buffer) were also included in SPR experiments as a baseline for double referencing. The BIAevaluation software (GE Healthcare) was used for the calculation of k_a, k_d, and K_D by curve fitting the data with a Langmuir 1:1 binding model.

6.7 Plasmid construction

Plasmids used in this study are described in Table below

Plasmid	Description	Source
pDD18	P_{BAD} promoter; pBR322 *ori*, Cm^r	Laboratory stock
pDD322-TatABC	*E. coli* TatABC under native promoter ;pBR322 *ori*, Tet^r	Laboratory stock
pDD18-spTorA-RGS-6xHis-pE59::MEK1R4F	pE59 and $MEK1^{R4F}$ with N-terminal spTorA signal and N-terminal RGS-6xHis tag in pDD18-Cm	This study
pDD18-spTorA-RGS-6xHis-pE59	pE59 with N-terminal spTorA signal and N-terminal RGS-6xHis tag in pDD18-Cm	This study
pDD18-spTorA-RGS-6xHis-Off7::MEK1R4F	Off7 and $MEK1^{R4F}$ with N-terminal spTorA signal and N-terminal RGS-6xHis tag in pDD18-Cm	This study
pDD18-spTorA-RGS-6xHis-Off7	Off7 with N-terminal spTorA signal and N-terminal RGS-6xHis tag in pDD18-Cm	This study
pDD322-TatABC-ERK2-Bla	ERK2-Bla with N-terminal FLAG in pDD322-TatABC	This study
pDD322-TatABC-mERK2-Bla	mERK2-Bla with N-terminal FLAG in pDD322-TatABC	This study

pDD322-TatABC-JNK2-Bla	JNK2-Bla with N-terminal FLAG in pDD322-TatABC	This study
pDST67	T5 promoter; Col E1 *ori*, Amp^r	Plückthun group
pDST67-RGS-6xHis-pE59	pE59 with N-terminal RGS-6xHis tag in pDST67	This study
pDST67-RGS-6xHis-pEM1	pEM1 with N-terminal RGS-6xHis tag in pDST67	This study
pDST67-RGS-6xHis-pEM2	pEM2 with N-terminal RGS-6xHis tag in pDST67	This study
pDST67-RGS-6xHis-pEM3	pEM3 with N-terminal RGS-6xHis tag in pDST67	This study
pLK1	P_1 promoter; Col E1 *ori*, Amp^r	Plückthun group
pLK1-Avi-ERK2-6xHis	ERK2 with N-terminal avi tag and C-terminal 6xHis tag	Plückthun group
pLK1-Avi-ERK2-6xHis:$MEK1^{R4F}$	ERK2 with N-terminal avi tag and C-terminal 6xHis tag and $MEK1^{R4F}$	Plückthun group

CHAPTER 4

FUTURE DIRECTIONS: ENHANCING PhLI-TRAP FOR UNCOVERED PRODUCTIVE BINDERS FOR POST-TRANSLATIONAL MODIFIED TARGETS

1. Introduction

So far, development of PhLI-TRAP, a novel platform technology for engineering and selection of binding proteins against phospho-modified target, in particular an oncogenic protein, ERK2, based on the bacterial Tat pathway has been discussed in previous chapters. The remarkable ability of the Tat pathway to transport non-native Tat substrate, including bacterial and eukaryotic proteins (ERK2), was demonstrated. As a proof-of-concept to be used as a protein selection technology, PhLI-TRAP was used to isolate DARPin variants towards pERK2 from site saturation mutagenesis libraries. All selected variants possess superior traits in terms of binding affinity than the parental pE59 DARPin by just one-round of mutagenesis. Most importantly, all variants selected from this system still retain their specificity against pERK2. This might be because the selection condition was performed in the presence of $MEK1^{R4F}$; consequently, cells were pressured to select only variants that bind exclusively to pERK2. Then, inspired by success in selection of DARPin variants with matured binding affinity, FLI-TRAP was then used to show it can be utilize to select binder with altered specificity. Similar method to previous library selection, this was simply performed by selection of a library pool in the absence of $MEK1^{R4F}$, and I selected variants that are able to accept ERK2 as a new target; however, these variants are promiscuous in binding to both ERK2 and pERK2. According to these findings, this method absolutely offers a fast and simple procedure that only involves

transformation of plasmids, functional expression of fusion proteins, and analysis of grown bacterial cells in a selective medium as a phenotypic readout. Moreover, as the intrinsic quality control of the Tat pathway was utilized in the selection method, it gives relatively low background of false-positive results. For instance, from 10 variants selected from a library pool with a purpose to select better binders, all performed better than the parental DARins. The method can be performed completely in vivo; therefore, it eliminates the requirement for purification and/or immobilization of a partner protein, thereby making this method a versatile technology. This chapter will discuss recommendations for further improvement on PhLI-TRAP.

2. High throughput PhLI-TRAP selection technology

All protein selection from library pools performed in this thesis is based on 2-plasmid system, that is, spTorA-RGS-6xHis-DARPin and $MEK1^{R4F}$ are in the same plasmid (pDD18) under the same L-arabinose inducing promoter, whilst ERK2-Bla is in another plasmid (pDD322) with constitutive promoter. With this strategy, after selection of potential variants against pERK2, and testing their binding affinity against pERK2, the specificity needs to be confirms by spot dilution experiment against ERK2. This involves subcloning of the sequence of each variant into pDD18 that is devoid of $MEK1^{R4F}$.

To expedite the process by elimination of cloning step, I hypothesize that by controlling expression of $MEK1^{R4F}$ by another tightly regulated promoter, we are able to control the functional expression of $MEK1^{R4F}$ by just adding a specific inducer. Then, with this strategy, the cloning step will be omitted and the binding specificity can be

simply performed by just single spotting of variants on selective media with and without a specific inducer.

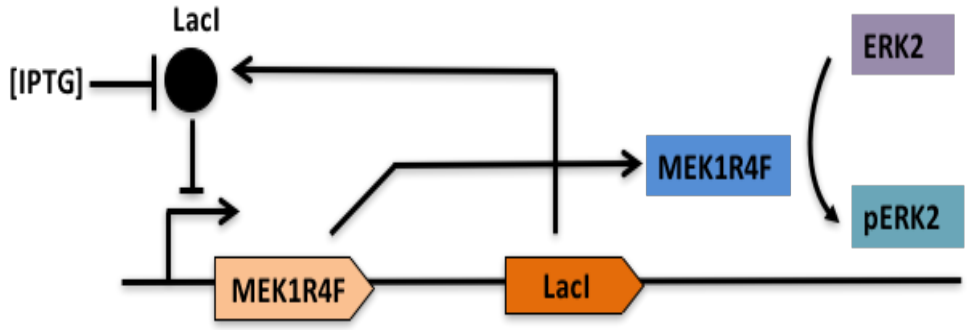


Figure 4.1. Schematic of a modified version of PhLI-TRAP by using IPTG as a toggle switch to control functional expression of MEK^{1R4F} in pEXT22

To preliminary test that this hypothesis holds true, MEK1^{R4F} was cloned in pEXT22 where the expression of kinase is controlled by Tac promoter using IPTG as an inducer. A schematic of this modified phospho-FLI-TRAP is shown in Fig. 4.1. When IPTG is added into a medium, the expression of MEK1R4F is turned on, thereby phosphorylation of ERK2 to be pERK2. As a proof of concept, spot dilution experiment (3-plasmid system) was performed at various concentrations of Carb with/without supplementation of IPTG, and the results were depicted in Fig. 4.2. Obviously, when IPTG was supplemented into media, antibiotic resistance was much higher than one without IPTG. To further validate this system, I plated E. coli cells harboring 3 plasmids (pDD18-spTorA-RGS-6xHis-pE59, pEXT22-MEK1^{R4F}, and pDD322-TatABC-ERK2-Bla) on a solid medium, and picked up some colonies for overnight culture. Next day, without any dilution step, 5 µl of the overnight culture was immediately spotted on a medium at different concentrations of Carb with/without IPTG. At 1600 mg/ml of Carb,

cell growth was observed on the medium with IPTG while no cell growth on one without IPTG as shown in Fig. 4.3.

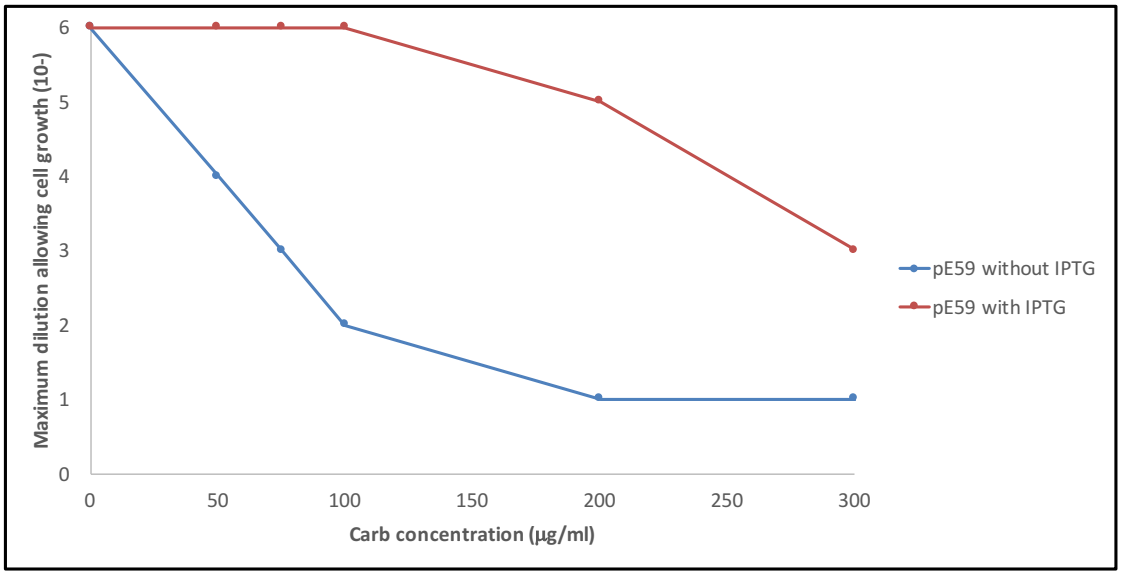


Figure 4.2. Spot dilution of *E. coli* cells with 3-plasmid system. The cells harbor pDD18-spTorA-RGS-6xHis-pE59, pEXT22-MEK1^{R4F}, and pDD322-TatABC-ERK2-Bla in the presence and absence of IPTG.

This finding strongly supports the notion that IPTG could be used as a toggle switch for MEK1^{R4F} functional expression, and in the future, this system could be used for protein selection by simply picking up colonies and performing just single spot on media with and without IPTG, all without the need of further cloning step.

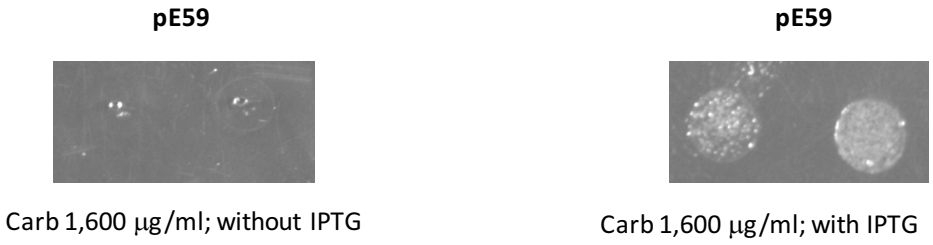


Figure 4.3. Single spot of overnight culture of *E. coli* cells. The cells harbor pDD18-spTorA-RGS-6xHis-pE59, pEXT22-MEK1^{R4F}, and pDD322-TatABC-ERK2-Bla on LBA with 1600 µg/ml of Carb in the presence and absence of IPTG.

3. Combination of PhLI-TRAP with ubiquibody technology

Selected variants from the affinity maturation library have shown to possess improved binding affinity and high selectivity against pERK2. Consequently, these selected variants will potentially be utilized to silence pERK2 expression in mammalian cells, in particular cancer cells. For therapeutic purpose, binding molecules not only should strong binding affinity and high fidelity to their target but also can eliminate the targets once it binds to. To generate such therapeutic proteins, DeLisa group has harnessed the cell's ubiquitin-proteasome pathway (UPP) guiding unneeded or damaged proteins to proteasomes, cellular complexes that break proteins apart as shown in Fig. 4.4. Our research group developed a novel approach called ubiquibodies. They are protein chimeras that combine the activity of E3 ubiquitin ligases with designer binding proteins (DBP), and they can usher selected proteins down the UPP for degradation. To knockdown phosphorylated ERK2 (pERK2), pERK2 ubiquibodies could be generated by simply fusing E3 ubiquitin ligases with pERK2 binding DARPin variants, including pEM1, pEM2, and pEM3 selected from PhLI-TRAP.

In addition to generation ubiquibodies by fusing PhLI-TRAP products to E3 ubiquitin ligases, we can improve the performance of ubiquibodies by using PhLI-TRAP technology. This can be accomplished by just cloning ubiquibodoy sequence after spTorA, making a library pool by error-prone mutagenesis or site-directed mutagenesis if desired, and performing selection of the library pool of ubiquibodies with the same method described in previous chapter. Ubiquibodies with desired traits, such as binding affinity, or stability could be then uncovered by this method.

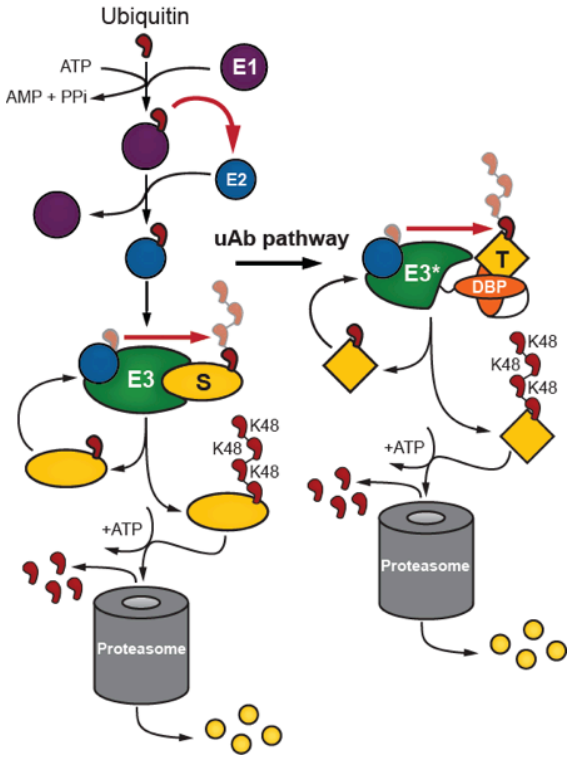


Figure 4.4. Redirecting the ubiquitin proteasome pathway. Schematic of redirecting the natural UPP with ubiquibodies (uAb). Naturally, the E1, E2 and E3 cascade tags substrate proteins (S) with polyubiquitin chains (K48 shown here) for proteasomal degradation. In the uAb pathway, the engineered E3*, where the natural substrate binding domain has been replaced with a designer binding protein (DBP), ubiquitinates the novel target (T) protein for proteasomal degradation [182].

4. Competitive PhLI-TRAP

It is clearly that all selected DARPin variants, especially pEM1, pEM2, and pEM3 from the library pool show stronger binding affinity and retain selectivity against pERK2. In another library with a purpose to reprogram specificity, EpEM6 and EpEM7 selected in the absence of kinase show their binding to ERK2; however, they also show promiscuity in ERK2 binding. Inspired by these findings of EpEM6 and EpEM7, a new version of PhLI-TRAP called competitive PhLI-TRAP is proposed here to attain the goal to obtain variants with desired high fidelity to target.

I hypothesize that by addition of a competitor into selection system, the competitor will enable more pressure to selection condition to obtain desired traits. In the event that we would like to select variants with higher binding affinity to pERK2, but not its counterpart, mutated ERK2 whose phosphorylation residues (Thr and Tyr) are mutated to Glu and Phe to prevent phosphorylation will be used as a competitor as shown in Fig. 4.5. In this system, variants that bind exclusively to pERK2 will be colocalized to periplasmic space, thus conferring antibiotic resistance, while ones with ERK2 binding or promiscuous binding will be retained in cytoplasm. With the competitor, it is highly possible that selected clones might specifically bind to pERK2 with high fidelity. Conversely, if variants with high fidelity to ERK2 are desired property, mutated ERK2-Bla will be used in the selection system, and ERK2 without Bla fusion which functions as a competitor will be coexpressed in the same cell. With this competitive PhLI-TRAP, variants with desired specificity will be plausibly selected with higher possibility.

In addition to selection of productive binders with high fidelity against either ERK2 or pERK2, competitive PhLI-TRAP could be potentially used for selection of binders to bind specifically to targets with high sequence homology, including ERK1/2 and JNK1/2. For instance, if ERK2 binder is desired, ERK1 will be used as a competitor and ERK2-Bla fusion protein will be used in the selection system.

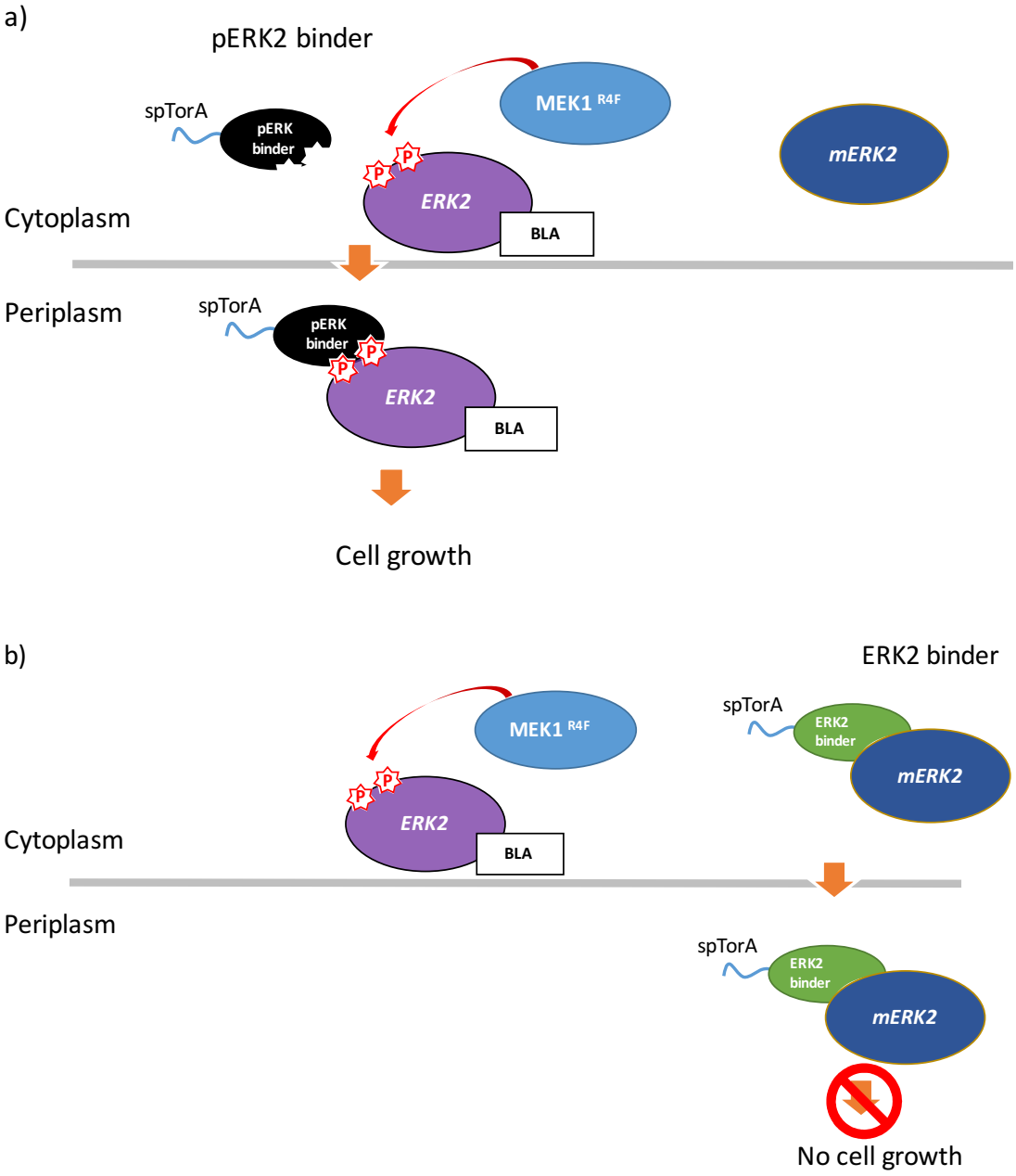


Figure 4.5. Schematic of competitive PhLI-TRAP to select productive binders with high binding affinity and fidelity to pERK2. Mutated ERK2 is coexpress and functions as a competitor. (a) If spTorA-DARPin binds specially to pERK2-Bla, protein complex will be colocalized to periplasm, thus allowing cell growth. (b) If spTorA-DARPin binds preferentially to ERK2, protein complex will be colocalized, but not allowing cell growth due to the lack of Bla fusion.

REFERENCES

1. Jensen, O. N. (2004). Modification-specific proteomics: characterization of post-translational modifications by mass spectrometry. *Current opinion in chemical biology*, *8*(1), 33-41.
2. Ayoubi, T. A., & Van De Ven, W. J. (1996). Regulation of gene expression by alternative promoters. *The FASEB Journal*, *10*(4), 453-460.
3. Duan, G., & Walther, D. (2015). The roles of post-translational modifications in the context of protein interaction networks. *PLoS computational biology*, *11*(2), e1004049.
4. Walsh, C. T., Garneau-Tsodikova, S., & Gatto, G. J. (2005). Protein posttranslational modifications: the chemistry of proteome diversifications. *Angewandte Chemie International Edition*, *44*(45), 7342-7372.
5. Liu, Z., Ma, Q., Cao, J., Gao, X., Ren, J., & Xue, Y. (2011). GPS-PUP: computational prediction of pupylation sites in prokaryotic proteins. *Molecular BioSystems*, *7*(10), 2737-2740.
6. Khoury, G. A., Baliban, R. C., & Floudas, C. A. (2011). Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database. *Scientific reports*, *1*, 90.
7. Cohen, P. (2000) The regulation of protein function by multisite phosphorylation--a 25 year update. *Trends in biochemical sciences* 25(12), 596-601.

8. Pinna, L. A., & Ruzzene, M. (1996). How do protein kinases recognize their substrates?. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, *1314*(3), 191-225.

9. Manning, G., Whyte, D. B., Martinez, R., Hunter, T., & Sudarsanam, S. (2002). The protein kinase complement of the human genome. *Science*, *298*(5600), 1912-1934.

10. Zhang, H., & Pelech, S. (2012) Using protein microarrays to study phosphorylation-mediated signal transduction. *Semin Cell Dev Biol,* 23(8), 872-882.

11. Russo, A.A., Jeffrey, P.D., & Pavletich, N.P. (1996) Structural basis of cyclin-dependent kinase activation by phosphorylation. *Nat Struct Biol* 3(8), 696-700.

12. Welburn, J. P., Tucker, J. A., Johnson, T., Lindert, L., Morgan, M., Willis, A., ... & Endicott, J. A. (2007). How tyrosine 15 phosphorylation inhibits the activity of cyclin-dependent kinase 2-cyclin A. *Journal of Biological Chemistry*, *282*(5), 3173-3181.

13. Canagarajah, B.J., Khokhlatchev, A., Cobb, M.H., & Goldsmith, E.J. (1997) Activation mechanism of the MAP kinase ERK2 by dual phosphorylation. *Cell,* 90(5), 859-869.

14. Ross, A. H., Baltimore, D., & Eisen, H. N. (1981). Phosphotyrosine-containing proteins isolated by affinity chromatography with antibodies to a synthetic hapten. *Nature*, *294*(5842), 654.

15. Nairn, A. C., Detre, J. A., Casnellie, J. E., & Greengard, P. (1982). Serum antibodies that distinguish between the phospho-and dephospho-forms of a phosphoprotein. *Nature*, *299*(5885), 734.

16. Dopfer, E. P., Schöpf, B., Louis-Dit-Sully, C., Dengler, E., Höhne, K., Kleščová, A., ... & Schamel, W. W. (2010). Analysis of novel phospho-ITAM specific antibodies in a S2 reconstitution system for TCR–CD3 signalling. *Immunology letters*, *130*(1-2), 43-50.

17. Zhou, H., Watts, J. D., & Aebersold, R. (2001). A systematic approach to the analysis of protein phosphorylation. *Nature biotechnology*, *19*(4), 375.

18. Witze, E.S., Old, W.M., Resing, K.A., & Ahn, N.G. (2007) Mapping protein post-translational modifications with mass spectrometry. *Nat Methods,* 4(10), 798-806.

19. Garcia, B.A., Shabanowitz, J., & Hunt, D.F. (2005) Analysis of protein phosphorylation by mass spectrometry. *Methods,* 35(3), 256-264.

20. Choudhary, C., & Mann, M. (2010) Decoding signalling networks by mass spectrometry-based proteomics. *Nat Rev Mol Cell Biol,* 11(6), 427-439.

21. Taussig, M. J., Stoevesandt, O., Borrebaeck, C. A., Bradbury, A. R., Cahill, D., Cambillau, C., ... & Gibson, T. J. (2007). ProteomeBinders: planning a European resource of affinity reagents for analysis of the human proteome. *Nature methods*, *4*(1), 13.

22. Bendall, S. C., Simonds, E. F., Qiu, P., El-ad, D. A., Krutzik, P. O., Finck, R., ... & Balderas, R. S. (2011). Single-cell mass cytometry of differential

immune and drug responses across a human hematopoietic continuum. *Science*, *332*(6030), 687-696.

23. Sachs, K., Perez, O., Pe'er, D., Lauffenburger, D.A., & Nolan, G.P. (2005) Causal protein-signaling networks derived from multiparameter single-cell data. *Science*, 308(5721), 523-529.

24. Perez, O.D. & Nolan, G.P. (2002) Simultaneous measurement of multiple active kinase states using polychromatic flow cytometry. *Nat Biotechnol,* 20(2), 155-162.

25. Pelech, S., Jelinkova, L., Susor, A., Zhang, H., Shi, X., Pavlok, A., ... & Kovarova, H. (2008). Antibody microarray analyses of signal transduction protein expression and phosphorylation during porcine oocyte maturation. *Journal of proteome research*, *7*(7), 2860-2871.

26. Haqshenas, G., Wu, J., Simpson, K. J., Daly, R. J., Netter, H. J., Baumert, T. F., & Doerig, C. (2017). Signalome-wide assessment of host cell response to hepatitis C virus. *Nature communications*, *8*, 15158.

27. Gembitsky, D.S., Lawlor, K., Jacovina, A., Yaneva, M., & Tempst, P. (2004). A prototype antibody microarray platform to monitor changes in protein tyrosine phosphorylation. *Mol Cell Proteomics,* 3(11), 1102-1118.

28. Brumbaugh, K., Johnson, W., Liao, W. C., Lin, M. S., Houchins, J. P., Cooper, J., ... & Campos-Gonzalez, R. (2011). Overview of the generation, validation, and application of phosphosite-specific antibodies. In *Signal Transduction Immunohistochemistry* (pp. 3-43). Humana Press.

29. Behring, EV. (1890) Studies on the development of diphtheria immunity in animals. Dtsch. Med. Wschr. 16, 1145–1147

30. Ehrlich, P. (1901). Die Seitenkettentheorie und ihre Gegner. Muenchner medizinische Wochenschrift, 2123-2124. 204.

31. Winau, F., Westphal, O., & Winau, R. (2004). Paul Ehrlich—in search of the magic bullet. *Microbes and Infection*, *6*(8), 786-789.

32. Rhorer, J., Ambrose, C. S., Dickinson, S., Hamilton, H., Oleka, N. A., Malinoski, F. J., & Wittes, J. (2009). Efficacy of live attenuated influenza vaccine in children: a meta-analysis of nine randomized clinical trials. *Vaccine*, *27*(7), 1101-1110.

33. Grimm, S. (2011). *Ribosome display for selection and evolution of affibody molecules* (Doctoral dissertation, KTH Royal Institute of Technology).

34. Lobato, M. N., & Rabbitts, T. H. (2003). Intracellular antibodies and challenges facing their use as therapeutic agents. *Trends in molecular medicine*, *9*(9), 390-396.

35. Gebauer, M., & Skerra, A. (2009). Engineered protein scaffolds as next-generation antibody therapeutics. *Curr Opin Chem Biol,* 13, 245-55.

36. Foote, J., & Winter, G. (1992). Antibody framework residues affecting the conformation of the hypervariable loops. *J Mol Biol,* 224, 487-99.

37. OGAWA, R., OHTSUKA, M., SASADAIRA, H., HIRASA, M., YABE, H., UCHIDA, H., & WATANABE, Y. (1985). Increase of phosphotyrosine-

containing proteins in human carcinomas. *Japanese Journal of Cancer Research GANN*, *76*(11), 1049-1055.

38. Dopfer, E. P., Schöpf, B., Louis-Dit-Sully, C., Dengler, E., Höhne, K., Kleščová, A., ... & Schamel, W. W. (2010). Analysis of novel phospho-ITAM specific antibodies in a S2 reconstitution system for TCR–CD3 signalling. *Immunology letters*, *130*(1-2), 43-50.

39. DiGiovanna, M.P., & Stern, D.F. (1995) Activation state-specific monoclonal antibody detects tyrosine phosphorylated p185neu/erbB-2 in a subset of human breast tumors overexpressing this receptor. *Cancer Res*, 55(9), 1946–1955.

40. Nita-Lazar, A., Saito-Benz, H., & White, F. M. (2008). Quantitative phosphoproteomics by mass spectrometry: past, present, and future. *Proteomics*, *8*(21), 4433-4443.

41. WoÈrn, A., & PluÈckthun, A. (2001). Stability engineering of antibody single-chain Fv fragments1. *Journal of molecular biology*, *305*(5), 989-1010.

42. Bendall, S. C., Simonds, E. F., Qiu, P., El-ad, D. A., Krutzik, P. O., Finck, R., ... & Balderas, R. S. (2011). Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum. *Science*, *332*(6030), 687-696.

43. Uhlen, M. & Hober, S. (2009). Generation and validation of affinity reagents on a proteome- wide level. *Journal of Molecular Recognition*, 22(2), 57-64.

44. Skerra, A., & Plückthun, A. (1988). Assembly of a functional immunoglobulin Fv fragment in Escherichia coli. *Science,* 240, 1038-41.

45. Bird, R. E., Hardman, K. D., Jacobson, J. W., Johnson, S., Kaufman, B. M., Lee, S. M., Lee, T., Pope, S. H., Riordan, G. S. & Whitlow, M. (1988). Single-chain antigen binding proteins. *Science,* 242, 423-6.

46. Better, M., Chang, C. P., Robinson, R. R. & Horwitz, A. H. (1988). Escherichia coli secretion of an active chimeric antibody fragment. *Science,* 240, 1041-3.

47. Holliger, P., & Hudson, P. J. (2005). Engineered antibody fragments and the rise of single domains. *Nature biotechnology*, 23(9), 1126.

48. Ghose, T., Cerini, M., Carter, M. & Nairn, R. C. (1967). Immunoradioactive agent against cancer. *Br Med J* 1, 90-3.

49. Muyldermans, S., Baral, T. N., Retamozzo, V. C., De Baetselier, P., De Genst, E., Kinne, J., Leonhardt, H., Magez, S., Nguyen, V. K., Revets, H., Rothbauer, U., Stijlemans, B., Tillib, S., Wernery, U., Wyns, L., Hassanzadeh-Ghassabeh, G. & Saerens, D. (2009). Camelid immunoglobulins and nanobody technology. *Vet Immunol Immunopathol,* 128, 178-83.

50. Pltickthun, A. (2009). Alternative scaffolds: expanding the options of antibodies. *Recombinant antibodies for immunotherapy*. *Cambridge University Press, Cambridge*, 243-271.

51. Boersma, Y. L., & Plückthun, A. (2011). DARPins and other repeat protein scaffolds: advances in engineering and applications. *Current opinion in biotechnology*, *22*(6), 849-857.

52. Stumpp, M. T., Binz, H. K., & Amstutz, P. (2008). DARPins: a new generation of protein therapeutics. *Drug discovery today*, *13*(15-16), 695-701.

53. Binz, H. K., Amstutz, P., & Plückthun, A. (2005). Engineering novel binding proteins from nonimmunoglobulin domains. *Nature biotechnology*, *23*(10), 1257.

54. Schlehuber, S., & Skerra, A. (2005). Anticalins as an alternative to antibody technology. *Expert opinion on biological therapy*, *5*(11), 1453-1462.

55. Flower, D. R., North, A. C. & Attwood, T. K. (1993). Structure and sequence relationships in the lipocalins and related proteins. *Protein Sci,* 2, 753-61.

56. Frand, A. R., Cuozzo, J. W. & Kaiser, C. A. (2000). Pathways for protein disulphide bond formation. *Trends Cell Biol,* 10, 203-10.

57. 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.

58. Mercader, J. V., & Skerra, A. (2002). Generation of anticalins with specificity for a nonsymmetric phthalic acid ester. *Analytical biochemistry*, *308*(2), 269-277.

59. Goetz, D. H., Willie, S. T., Armen, R. S., Bratt, T., Borregaard, N., & Strong, R. K. (2000). Ligand preference inferred from the structure of neutrophil gelatinase associated lipocalin. *Biochemistry*, *39*(8), 1935-1941.

60. Schlapschy, M., Grimm, S. & Skerra, A. (2006). A system for concomitant overexpression of four periplasmic folding catalysts to improve secretor protein production in Escherichia coli. *Protein Eng Des Sel,* 19, 385-90.

61. Nord, K., Nilsson, J., Nilsson, B., Uhlen, M. & Nygren, P. A. (1995). A combinatorial library of an alpha-helical bacterial receptor domain. Protein Eng Des Sel, 8, 601-8.

62. Foster, T. J. (2005). Immune evasion by staphylococci. *Nat Rev Microbiol,* 3, 948-58.

63. Löfblom, J., Feldwisch, J., Tolmachev, V., Carlsson, J., Ståhl, S., & Frejd, F. Y. (2010). Affibody molecules: engineered proteins for therapeutic, diagnostic and biotechnological applications. *FEBS letters*, 584(12), 2670-2680.

64. Nilsson, B., Moks, T., Jansson, B., Abrahmsen, L., Elmblad, A., Holmgren, E., Henrichson, C., Jones, T. A. & Uhlen, M. (1987). A synthetic IgG-binding domain based on staphylococcal protein A. *Protein Eng Des Sel*, 1, 107-13.

65. Nygren, P.-Å. (2008). Alternative binding proteins: Affibody binding proteins developed from a small three-helix bundle scaffold. *FEBS J*, 275(11), 2668-2676.

66. Arora, P., Oas, T. G. & Myers, J. K. (2004). Fast and faster: a designed variant of the Bdomain of protein A folds in 3 microsec. *Protein Sci,* 13(4), 847-53.

67. Feldwisch, J., Tolmachev, V., Lendel, C., Herne, N., Sjoberg, A., Larsson, B., Rosik, D., Lindqvist, E., Fant, G., Hoiden-Guthenberg, I., Galli, J., Jonasson, P. & Abrahmsen, L. (2010). Design of an optimized scaffold for affibody molecules. *J Mol Biol*, 398(2), 232-47.

68. Kronqvist, N., Malm, M., Göstring, L., Gunneriusson, E., Nilsson, M., Höidén Guthenberg, I., ... & Löfblom, J. (2010). Combining phage and staphylococcal surface display for generation of ErbB3-specific Affibody molecules. *Protein Engineering, Design & Selection*, *24*(4), 385-396.

69. Grimm, S., Yu, F., & Nygren, P. Å. (2011). Ribosome display selection of a murine IgG 1 Fab binding affibody molecule allowing species selective recovery of monoclonal antibodies. *Molecular biotechnology*, *48*(3), 263-276.

70. Gunneriusson, E., Nord, K., Uhlen, M., & Nygren, P. Å. (1999). Affinity maturation of a Taq DNA polymerase specific affibody by helix shuffling. *Protein Engineering*, *12*(10), 873-878.

71. Grönwall, C., Sjöberg, A., Ramström, M., Höidén-Guthenberg, I., Hober, S., Jonasson, P., & Ståhl, S. (2007). Affibody-mediated transferrin depletion for proteomics applications. *Biotechnology journal*, *2*(11), 1389-1398.

72. Löfdahl, P. Å., Nord, O., Janzon, L., & Nygren, P. Å. (2009). Selection of TNF-α binding affibody molecules using a β-lactamase protein fragment complementation assay. *New biotechnology*, *26*(5), 251-259.

73. Wikman, M., Steffen, A. C., Gunneriusson, E., Tolmachev, V., Adams, G. P., Carlsson, J., & Ståhl, S. (2004). Selection and characterization of HER2/neu-binding affibody ligands. *Protein Engineering Design and Selection*, *17*(5), 455-462.

74. Li, J., Lundberg, E., Vernet, E., Larsson, B., Höidén-Guthenberg, I., & Gräslund, T. (2010). Selection of affibody molecules to the ligand-binding site of the insulin-like growth factor-1 receptor. *Biotechnology and applied biochemistry*, *55*(2), 99-109.

75. Eklund, M., Axelsson, L., Uhlén, M., & Nygren, P. Å. (2002). Anti-idiotypic protein domains selected from protein A-based affibody libraries. *Proteins: Structure, Function, and Bioinformatics*, *48*(3), 454-462.

76. Wållberg, H., Löfdahl, P. Å., Tschapalda, K., Uhlén, M., Tolmachev, V., Nygren, P. Å., & Ståhl, S. (2011). Affinity recovery of eight HER2-binding affibody variants using an anti-idiotypic affibody molecule as capture ligand. *Protein expression and purification*, *76*(1), 127-135.

77. Rönnmark, J., Grönlund, H., Uhlén, M., & Nygren, P. Å. (2002). Human immunoglobulin A (IgA)-specific ligands from combinatorial engineering of protein A. *The FEBS Journal*, *269*(11), 2647-2655.

78. Wikman, M., Rowcliffe, E., Friedman, M., Henning, P., Lindholm, L., Olofsson, S., & Ståhl, S. (2006). Selection and characterization of an HIV-1 gp120-binding affibody ligand. *Biotechnology and applied biochemistry*, *45*(2), 93-105.

79. Sandström, K., Xu, Z., Forsberg, G., & Nygren, P. Å. (2003). Inhibition of the CD28–CD80 co-stimulation signal by a CD28-binding affibody ligand developed by combinatorial protein engineering. *Protein engineering*, *16*(9), 691-697.

80. Grönwall, C., Jonsson, A., Lindström, S., Gunneriusson, E., Ståhl, S., & Herne, N. (2007). Selection and characterization of Affibody ligands binding to Alzheimer amyloid β peptides. *Journal of biotechnology*, *128*(1), 162-183.

81. Orlova, A., Magnusson, M., Eriksson, T. L., Nilsson, M., Larsson, B., Höidén-Guthenberg, I., ... & Nilsson, F. Y. (2006). Tumor imaging using a picomolar affinity HER2 binding affibody molecule. *Cancer research*, *66*(8), 4339-4348.

82. Baum, R. P., Prasad, V., Müller, D., Schuchardt, C., Orlova, A., Wennborg, A., ... & Feldwisch, J. (2010). Molecular imaging of HER2-expressing malignant tumors in breast cancer patients using synthetic 111In-or 68Ga-labeled affibody molecules. *Journal of nuclear medicine*, *51*(6), 892.

83. Baum, R. P., Prasad, V., Müller, D., Schuchardt, C., Orlova, A., Wennborg, A., ... & Feldwisch, J. (2010). Molecular imaging of HER2-expressing

malignant tumors in breast cancer patients using synthetic 111In-or 68Ga-labeled affibody molecules. *Journal of nuclear medicine*, *51*(6), 892.

84. Zielinski, R., Lyakhov, I., Jacobs, A., Chertov, O., Kramer-Marek, G., Francella, N., ... & Capala, J. (2009). Affitoxin–a novel recombinant, HER2-specific, anti-cancer agent for targeted therapy of HER2-positive tumors. *Journal of immunotherapy (Hagerstown, Md.: 1997)*, *32*(8), 817.

85. Lipovšek, D. (2010). Adnectins: engineered target-binding protein therapeutics. *Protein Engineering, Design & Selection*, *24*(1-2), 3-9.

86. Pankov, R., & Yamada, K. M. (2002). Fibronectin at a glance. *Journal of cell science*, *115*(20), 3861-3863.

87. Plaxco, K. W., Spitzfaden, C., Campbell, I. D. & Dobson, C. M. (1996). Rapid refolding of a proline-rich all-beta-sheet fibronectin type III module. *Proc Natl Acad Sci USA*, 93, 10703-6.

88. Koide, A., Bailey, C. W., Huang, X., & Koide, S. (1998). The fibronectin type III domain as a scaffold for novel binding proteins1. *Journal of molecular biology*, *284*(4), 1141-1151.

89. Hackel, B. J., Kapila, A., & Wittrup, K. D. (2008). Picomolar affinity fibronectin domains engineered utilizing loop length diversity, recursive mutagenesis, and loop shuffling. *Journal of molecular biology*, *381*(5), 1238-1252.

90. Plückthun, A. (2015). Designed ankyrin repeat proteins (DARPins): binding proteins for research, diagnostics, and therapy. *Annual review of pharmacology and toxicology*, *55*, 489-511.

91. Pancer, Z., Amemiya, C. T., Ehrhardt, G. R., Ceitlin, J., Gartland, G. L., & Cooper, M. D. (2004). Somatic diversification of variable lymphocyte receptors in the agnathan sea lamprey. *Nature*, *430*(6996), 174.

92. Herrin, B. R., Alder, M. N., Roux, K. H., Sina, C., Ehrhardt, G. R., Boydston, J. A., ... & Cooper, M. D. (2008). Structure and specificity of lamprey monoclonal antibodies. *Proceedings of the National Academy of Sciences*, *105*(6), 2040-2045.

93. Hussain, S., Plückthun, A., Allen, T. M., & Zangemeister-Wittke, U. (2006). Chemosensitization of carcinoma cells using epithelial cell adhesion molecule–targeted liposomal antisense against bcl-2/bcl-xL. *Molecular cancer therapeutics*, *5*(12), 3170-3180.

94. Hussain, S., Plückthun, A., Allen, T. M., & Zangemeister-Wittke, U. (2007). Antitumor activity of an epithelial cell adhesion molecule–targeted nanovesicular drug delivery system. *Molecular cancer therapeutics*, *6*(11), 3019-3027.

95. Martin-Killias, P., Stefan, N., Rothschild, S., Plückthun, A., & Zangemeister-Wittke, U. (2011). A novel fusion toxin derived from an EpCAM-specific designed ankyrin repeat protein has potent antitumor activity. *Clinical cancer research*, *17*(1), 100-110.

96. Alley, S. C., Okeley, N. M., & Senter, P. D. (2010). Antibody–drug conjugates: targeted drug delivery for cancer. *Current opinion in chemical biology*, *14*(4), 529-537.

97. Winkler, J., Martin-Killias, P., Plückthun, A., & Zangemeister-Wittke, U. (2009). EpCAM-targeted delivery of nanocomplexed siRNA to tumor cells with designed ankyrin repeat proteins. *Molecular cancer therapeutics*, *8*(9), 2674-2683.

98. Steiner, D., Forrer, P., & Plückthun, A. (2008). Efficient selection of DARPins with sub-nanomolar affinities using SRP phage display. *Journal of molecular biology*, *382*(5), 1211-1227.

99. Kummer, L., Parizek, P., Rube, P., Millgramm, B., Prinz, A., Mittl, P. R., ... & Plückthun, A. (2012). Structural and functional analysis of phosphorylation-specific binders of the kinase ERK from designed ankyrin repeat protein libraries. *Proceedings of the National Academy of Sciences*, *109*(34), E2248-E2257.

100. Hoess, R. H. (2001). Protein design and phage display. *Chemical Reviews*, *101*(10), 3205-3218.

101. Smith, G. P. (1985). Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science*, *228*(4705), 1315-1317.

102. Bass, S., Greene, R., & Wells, J. A. (1990). Hormone phage: an enrichment method for variant proteins with altered binding properties. *Proteins: Structure, Function, and Bioinformatics*, *8*(4), 309-314.

103. McCafferty, J., Griffiths, A. D., Winter, G., & Chiswell, D. J. (1990). Phage antibodies: filamentous phage displaying antibody variable domains. *nature*, *348*(6301), 552.

104. Clackson, T., Hoogenboom, H. R., Griffiths, A. D., & Winter, G. (1991). Making antibody fragments using phage display libraries. *Nature*, *352*(6336), 624.

105. Hemminki, A., Niemi, S., Hoffren, A. M., Hakalahti, L., Söderlund, H., & Takkinen, K. (1998). Specificity improvement of a recombinant anti-testosterone Fab fragment by CDRIII mutagenesis and phage display selection. *Protein engineering*, *11*(4), 311-319.

106. Pasqualini, R., & Ruoslahti, E. (1996). Organ targeting in vivo using phage display peptide libraries. *Nature*, *380*(6572), 364.

107. Joyce, J. A., Laakkonen, P., Bernasconi, M., Bergers, G., Ruoslahti, E., & Hanahan, D. (2003). Stage-specific vascular markers revealed by phage display in a mouse model of pancreatic islet tumorigenesis. *Cancer cell*, *4*(5), 393-403.

108. Bradbury, A. R., & Marks, J. D. (2004). Antibodies from phage antibody libraries. *Journal of immunological methods*, *290*(1-2), 29-49.

109. SMITH, G., & Scott, J. K. (1995). Libraries of peptides and proteins displayed on filamentous phage. In *Recombinant DNA Methodology II* (pp. 333-362).

110. Iannolo, G., Minenkova, O., Petruzzelli, R., & Cesareni, G. (1995). Modifying filamentous phage capsid: limits in the size of the major capsid protein. *Journal of molecular biology*, *248*(4), 835-844.

111. Malik, P., Terry, T. D., Gowda, L. R., Langara, A., Petukhov, S. A., Symmons, M. F., ... & Perham, R. N. (1996). Role of capsid structure and membrane protein processing in determining the size and copy number of peptides displayed on the major coat protein of filamentous bacteriophage. *Journal of molecular biology*, *260*(1), 9-21.

112. Steiner, D., Forrer, P., Stumpp, M. T., & Plückthun, A. (2006). Signal sequences directing cotranslational translocation expand the range of proteins amenable to phage display. *Nature biotechnology*, *24*(7), 823.

113. Thie, H., Schirrmann, T., Paschke, M., Dübel, S., & Hust, M. (2008). SRP and Sec pathway leader peptides for antibody phage display and antibody fragment production in E. coli. *New biotechnology*, *25*(1), 49-54.

114. Paschke, M., & Höhne, W. (2005). A twin-arginine translocation (Tat)-mediated phage display system. *Gene*, *350*(1), 79-88.

115. Hanes, J., & Plückthun, A. (1997). In vitro selection and evolution of functional proteins by using ribosome display. *Proceedings of the National Academy of Sciences*, *94*(10), 4937-4942.

116. Francisco, J. A., Earhart, C. F., & Georgiou, G. (1992). Transport and anchoring of beta-lactamase to the external surface of Escherichia coli. *Proceedings of the National Academy of Sciences*, *89*(7), 2713-2717.

117. Francisco, J. A., Campbell, R., Iverson, B. L., & Georgiou, G. (1993). Production and fluorescence-activated cell sorting of Escherichia coli expressing a functional antibody fragment on the external surface. *Proceedings of the National Academy of Sciences*, *90*(22), 10444-10448.

118. Brown, S. (1992). Engineered iron oxide-adhesion mutants of the Escherichia coli phage lambda receptor. *Proceedings of the National Academy of Sciences*, *89*(18), 8651-8655.

119. Lu, Z., Murray, K. S., Van Cleave, V., LaVallie, E. R., Stahl, M. L., & McCoy, J. M. (1995). Expression of thioredoxin random peptide libraries on the Escherichia coli cell surface as functional fusions to flagellin: a system designed for exploring protein-protein interactions. *Nature biotechnology*, *13*(4), 366.

120. Daugherty, P. S., Chen, G., Olsen, M. J., Iverson, B. L., & Georgiou, G. (1998). Antibody affinity maturation using bacterial surface display. *Protein engineering*, *11*(9), 825-832.

121. Mazor, Y., Van Blarcom, T., Mabry, R., Iverson, B. L., & Georgiou, G. (2007). Isolation of engineered, full-length antibodies from libraries expressed in Escherichia coli. *Nature biotechnology*, *25*(5), 563.

122. Georgiou, G., Stathopoulos, C., Daugherty, P. S., Nayak, A. R., Iverson, B. L., & Curtiss III, R. (1997). Display of heterologous proteins on the surface

of microorganisms: from the screening of combinatorial libraries to live recombinant vaccines. *Nature biotechnology*, *15*(1), 29-34.

123. Daugherty, P. S. (2007). Protein engineering with bacterial display. *Current opinion in structural biology*, *17*(4), 474-480.

124. Gibbs, K. A., Isaac, D. D., Xu, J., Hendrix, R. W., Silhavy, T. J., & Theriot, J. A. (2004). Complex spatial distribution and dynamics of an abundant Escherichia coli outer membrane protein, LamB. *Molecular microbiology*, *53*(6), 1771-1783.

125. Binder, U., Matschiner, G., Theobald, I., & Skerra, A. (2010). High-throughput sorting of an Anticalin library via EspP-mediated functional display on the Escherichia coli cell surface. *Journal of molecular biology*, *400*(4), 783-802.

126. Boder, E. T., & Wittrup, K. D. (1997). Yeast surface display for screening combinatorial polypeptide libraries. *Nature biotechnology*, *15*(6), 553.

127. Gai, S. A., & Wittrup, K. D. (2007). Yeast surface display for protein engineering and characterization. *Current opinion in structural biology*, *17*(4), 467-473.

128. Boder, E. T., Midelfort, K. S., & Wittrup, K. D. (2000). Directed evolution of antibody fragments with monovalent femtomolar antigen-binding affinity. *Proceedings of the National Academy of Sciences*, *97*(20), 10701-10705.

129. Garcia-Rodriguez, C., Levy, R., Arndt, J. W., Forsyth, C. M., Razai, A., Lou, J., Geren, I., Stevens, R. C. & Marks, J. D. (2007). Molecular evolution of antibody crossreactivity for two subtypes of type A botulinum neurotoxin. *Nat Biotechnol* 25(1), 107-16.

130. Wernérus, H., Lehtiö, J., Teeri, T., Nygren, P. Å., & Ståhl, S. (2001). Generation of metal-binding staphylococci through surface display of combinatorially engineered cellulose-binding domains. *Applied and environmental microbiology*, *67*(10), 4678-4684.

131. Kronqvist, N., Löfblom, J., Jonsson, A., Wernérus, H., & Ståhl, S. (2008). A novel affinity protein selection system based on staphylococcal cell surface display and flow cytometry. *Protein Engineering, Design & Selection*, *21*(4), 247-255.

132. Rockberg, J., Löfblom, J., Hjelm, B., Uhlén, M., & Ståhl, S. (2008). Epitope mapping of antibodies using bacterial surface display. *Nature Methods*, *5*(12), 1039.

133. Low, N. M., Holliger, P., & Winter, G. (1996). Mimicking somatic hypermutation: affinity maturation of antibodies displayed on bacteriophage using a bacterial mutator strain. *Journal of molecular biology*, *260*(3), 359-368.

134. Mattheakis, L. C., Bhatt, R. R., & Dower, W. J. (1994). An in vitro polysome display system for identifying ligands from very large peptide

libraries. *Proceedings of the National Academy of Sciences*, *91*(19), 9022-9026.

135. Makeyev, E. V., Kolb, V. A., & Spirin, A. S. (1996). Enzymatic activity of the ribosome- bound nascent polypeptide. *FEBS letters*, *378*(2), 166-170.

136. Osada, E., Shimizu, Y., Akbar, B. K., Kanamori, T., & Ueda, T. (2009). Epitope mapping using ribosome display in a reconstituted cell-free protein synthesis system. *Journal of biochemistry*, *145*(5), 693-700.

137. Ohashi, H., Shimizu, Y., Ying, B. W., & Ueda, T. (2007). Efficient protein selection based on ribosome display system with purified components. *Biochemical and biophysical research communications*, *352*(1), 270-276.

138. Matsuura, T., Yanagida, H., Ushioda, J., Urabe, I., & Yomo, T. (2007). Nascent chain, mRNA, and ribosome complexes generated by a pure translation system. *Biochemical and biophysical research communications*, *352*(2), 372-377.

139. Chodorge, M., Fourage, L., Ravot, G., Jermutus, L., & Minter, R. (2008). In vitro DNA recombination by L-Shuffling during ribosome display affinity maturation of an anti-Fas antibody increases the population of improved variants. *Protein Engineering, Design & Selection*, *21*(5), 343-351.

140. Roberts, R. W., & Szostak, J. W. (1997). RNA-peptide fusions for the in vitro selection of peptides and proteins. *Proceedings of the National Academy of Sciences*, *94*(23), 12297-12302.

141. Galán, A., Comor, L., Horvatić, A., Kuleš, J., Guillemin, N., Mrljak, V., & Bhide, M. (2016). Library-based display technologies: where do we stand?. *Molecular BioSystems*, *12*(8), 2342-2358.

142. Takahashi, T. T., Austin, R. J., & Roberts, R. W. (2003). mRNA display: ligand discovery, interaction analysis and beyond. *Trends in biochemical sciences*, *28*(3), 159-165.

143. Seelig, B., & Szostak, J. W. (2007). Selection and evolution of enzymes from a partially randomized non-catalytic scaffold. *Nature*, *448*(7155), 828.

144. Plückthun, A. (2012). Ribosome display: a perspective. In *Ribosome Display and Related Technologies* (pp. 3-28). Springer, New York, NY.

145. Ullman, C. G., Frigotto, L., & Cooley, R. N. (2011). In vitro methods for peptide display and their applications. *Briefings in functional genomics*, *10*(3), 125-134.

146. Odegrip, R., Coomber, D., Eldridge, B., Hederer, R., Kuhlman, P. A., Ullman, C., ... & McGregor, D. (2004). CIS display: in vitro selection of peptides from libraries of protein–DNA complexes. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(9), 2806-2810.

147. Reichert, J. M., Jacob, N. M., & Amanullah, A. (2009). Second international conference on accelerating biopharmaceutical development. *Biotechnology progress*, *25*(4), 1198-1215.

148. Reiersen, H., Løbersli, I., Løset, G. Å., Hvattum, E., Simonsen, B., Stacy, J. E., ... & Marvik, O. J. (2005). Covalent antibody display—an in vitro antibody-DNA library selection system. *Nucleic acids research*, *33*(1), e10-e10.

149. Howorka, S. (Ed.). (2011). *Molecular assembly in natural and engineered systems* (Vol. 103). Academic Press.

150. Jackrel, M. E., Cortajarena, A. L., Liu, T. Y., & Regan, L. (2010). Screening libraries to identify proteins with desired binding activities using a split-GFP reassembly assay. *ACS chemical biology*, *5*(6), 553-562.

151. Löfdahl, P. Å., Nord, O., Janzon, L., & Nygren, P. Å. (2009). Selection of TNF-α binding affibody molecules using a β-lactamase protein fragment complementation assay. *New biotechnology*, *26*(5), 251-259.

152. MoÈssner, E., Koch, H., & PluÈckthun, A. (2001). Fast selection of antibodies without antigen purification: adaptation of the protein fragment complementation assay to select antigen-antibody pairs1. *Journal of molecular biology*, *308*(2), 115-122.

153. Waraho, D., & DeLisa, M. P. (2012). Identifying and Optimizing Intracellular Protein–Protein Interactions Using Bacterial Genetic Selection. In *Synthetic Gene Networks* (pp. 125-143). Humana Press.

154. Robinson, C., & Bolhuis, A. (2004). Tat-dependent protein targeting in prokaryotes and chloroplasts. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, *1694*(1-3), 135-147.

155. Robinson, C., & Bolhuis, A. (2001). Protein targeting by the twin-arginine translocation pathway. *Nature Reviews Molecular Cell Biology*, *2*(5), 350.

156. Müller, M., & Bernd Klösgen, R. (2005). The Tat pathway in bacteria and chloroplasts. *Molecular membrane biology*, *22*(1-2), 113-121.

157. Müller, M. (2005). Twin-arginine-specific protein export in Escherichia coli. *Research in microbiology*, *156*(2), 131-136.

158. Lee, P. A., Tullman-Ercek, D., & Georgiou, G. (2006). The bacterial twin-arginine translocation pathway. *Annu. Rev. Microbiol.*, *60*, 373-395.

159. Weiner, J. H., Bilous, P. T., Shaw, G. M., Lubitz, S. P., Frost, L., Thomas, G. H., ... & Turner, R. J. (1998). A novel and ubiquitous system for membrane targeting and secretion of cofactor-containing proteins. *Cell*, *93*(1), 93-101.

160. Berks, B. C. (1996). A common export pathway for proteins binding complex redox cofactors?. *Molecular microbiology*, *22*(3), 393-404.

161. Berks, B. C., Sargent, F., De Leeuw, E., Hinsley, A. P., Stanley, N. R., Jack, R. L., ... & Palmer, T. (2000). A novel protein transport system involved in the biogenesis of bacterial electron transfer chains. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, *1459*(2-3), 325-330.

162. Berks, B. C., Sargent, F., & Palmer, T. (2000). The Tat protein export pathway. *Molecular microbiology*, *35*(2), 260-274.

163. Ribnicky, B., Van Blarcom, T., & Georgiou, G. (2007). A scFv antibody mutant isolated in a genetic screen for improved export via the twin arginine transporter pathway exhibits faster folding. *Journal of molecular biology*, *369*(3), 631-639.

164. Rodrigue, A., Chanal, A., Beck, K., Müller, M., & Wu, L. F. (1999). Co-translocation of a periplasmic enzyme complex by a hitchhiker mechanism through the bacterial tat pathway. *Journal of Biological Chemistry*, *274*(19), 13223-13228.

165. Waraho, D., & DeLisa, M. P. (2009). Versatile selection technology for intracellular protein–protein interactions mediated by a unique bacterial hitchhiker transport mechanism. *Proceedings of the National Academy of Sciences*, *106*(10), 3692-3697.

166. Bogsch, E. G., Sargent, F., Stanley, N. R., Berks, B. C., Robinson, C., & Palmer, T. (1998). An essential component of a novel bacterial protein export system with homologues in plastids and mitochondria. *Journal of Biological Chemistry*, *273*(29), 18003-18006.

167. Sargent, F., Bogsch, E. G., Stanley, N. R., Wexler, M., Robinson, C., Berks, B. C., & Palmer, T. (1998). Overlapping functions of components of a bacterial Sec- independent protein export pathway. *The EMBO Journal*, *17*(13), 3640-3650.

168. Settles, A. M., Yonetani, A., Baron, A., Bush, D. R., Cline, K., & Martienssen, R. (1997). Sec-independent protein translocation by the maize Hcf106 protein. *Science*, *278*(5342), 1467-1470.

169. Behrendt, J., Standar, K., Lindenstrauß, U., & Brüser, T. (2006). Topological studies on the twin-arginine translocase component TatC. *FEMS microbiology letters*, *234*(2), 303-308.

170. Mori, H., & Cline, K. (2002). A twin arginine signal peptide and the pH gradient trigger reversible assembly of the thylakoid ΔpH/Tat translocase. *The Journal of cell biology*, *157*(2), 205-210.

171. Cline, K., & McCaffery, M. (2007). Evidence for a dynamic and transient pathway through the TAT protein transport machinery. *The EMBO journal*, *26*(13), 3039-3049.

172. Gohlke, U., Pullan, L., McDevitt, C. A., Porcelli, I., de Leeuw, E., Palmer, T., ... & Berks, B. C. (2005). The TatA component of the twin-arginine protein transport system forms channel complexes of variable diameter. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(30), 10482-10486.

173. DeLisa, M. P., Tullman, D., & Georgiou, G. (2003). Folding quality control in the export of proteins by the bacterial twin-arginine translocation pathway. *Proceedings of the National Academy of Sciences*, *100*(10), 6115-6120.

174. Fisher, A. C., & DeLisa, M. P. (2009). Efficient isolation of soluble intracellular single-chain antibodies using the twin-arginine translocation machinery. *Journal of molecular biology*, *385*(1), 299-311.

175. Fisher, A. C., Kim, J. Y., Perez-Rodriguez, R., Tullman-Ercek, D., Fish, W., Henderson, L. A., & DeLisa, M. P. (2008). Exploration of twin-arginine translocation for the expression and purification of correctly folded proteins in Escherichia coli, Microbial Biotechnol 1, 403–415. *CrossRef Google Scholar*.

176. Fisher, A. C., Kim, W., & Delisa, M. P. (2006). Genetic selection for protein solubility enabled by the folding quality control feature of the twin- arginine translocation pathway. *Protein Science*, *15*(3), 449-458.

177. Feilmeier, B. J., Iseminger, G., Schroeder, D., Webber, H., & Phillips, G. J. (2000). Green fluorescent protein functions as a reporter for protein localization in Escherichia coli. *Journal of bacteriology*, *182*(14), 4068-4076.

178. Mansell, T. J., Linderman, S. W., Fisher, A. C., & DeLisa, M. P. (2010). A rapid protein folding assay for the bacterial periplasm. *Protein Science*, *19*(5), 1079-1090.

179. Jost, C., & Plückthun, A. (2014). Engineered proteins with desired specificity: DARPins, other alternative scaffolds and bispecific IgGs. *Current opinion in structural biology*, *27*, 102-112.

180. Varadarajan, N., Gam, J., Olsen, M. J., Georgiou, G., & Iverson, B. L. (2005). Engineering of protease variants exhibiting high catalytic activity and exquisite substrate selectivity. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(19), 6855-6860.

181. Waraho-Zhmayev, D., Meksiriporn, B., Portnoff, A. D., & DeLisa, M. P. (2014). Optimizing recombinant antibodies for intracellular function using hitchhiker-mediated survival selection. *Protein Engineering, Design & Selection*, *27*(10), 351-358.

182. Portnoff, A. D., Stephens, E. A., Varner, J. D., & DeLisa, M. P. (2014). Ubiquibodies, synthetic E3 ubiquitin ligases endowed with unnatural substrate specificity for targeted protein silencing. *Journal of Biological Chemistry*, *289*(11), 7844-7855.