Construction and functional validation of combinatorial designed ankyrin repeat protein (DARPin) libraries for phage display

Master thesis

University of Turku

Department of Life Technologies

Molecular Biotechnology and Diagnostics

May 2023

Matias Lappalainen

Abstract

Designed ankyrin repeat protein (DARPin) is one of the most advanced alternative scaffold proteins. In this study, a combinatorial DARPin library for phage display was constructed. Firstly, a binary DARPin phage display library (termed BiLiBST) was constructed containing variating tyrosine and serine residues. Five PelB signal sequence mutants were then analyzed to find the one with the best display levels for BiLiBST. These variants were found in earlier study on anti-GFP DARPin display using modified signal sequence, but it was not known if the increased display levels were a generic behavior irrespective of displaying either fixed DARPin protein or library DARPin. The capability of each PelB variant to display BiLiBST was determined by anti-DARPin immunoassay, and the phage stocks titers were normalized by total phage immunoassay. All tested PelB variants exhibited statistically significant improvement in display of the BiLiBST compared to the parental PelB and few variants had even twofold higher display efficiency than DsbA, which has previously been the standard for filamentous phage display of DARPins. The BiLiBST library was displayed with DN5 signal sequence and preselected for open reading frames by panning two rounds against anti-DARPin Fabs. The DARPin libraries were effectively purified from frameshifts as the frequency of frameshift decreased from 50 % to 13 %. The main DARPin library construction was performed by assembly PCR using purified BiLiBST as template and its validity was confirmed by selecting against the GST-tagged protein N from SARS-COV-2 for three rounds. Successful enrichment of binders was confirmed by phage immunoreactivity assay with signal to background ratios up to 11-fold. This is a solid foundation for further molecular diversification and library construction.

Contents

Abbreviations

CDR	Complementarity-determining region
DARPin	Designed ankyrin repeat protein
EGFR	Epidermal growth factor receptor
Fab	Fragment antigen binding
FACS	Fluorescence-activated cell sorting
FAP	Fluorogen-activated protein
HER2	Human epidermal growth factor receptor 2
HSA	Human serum albumin
Kd	Dissociation constant
OCCA	Oligonucleotide-directed chelate complementation assay
PCR	Polymerase chain reaction
RBD	Receptor-binding domain
ScFv	Single-chain variable fragment

1. Introduction

1.1 Properties of DARPins

Combinatorial protein libraries are a versatile source of bioreagents for diagnostics and therapy. Currently, many applications rely solely on antibodies, but in recent years these have been challenged by alternative scaffold proteins. By switching to synthetic binders built on a simple scaffold, the limitations found in antibodies can be negated. Limitations associated with antibodies are, for example, large molecule size, intradomain and interchain disulfide bond dependent stability which stops the folding in reducing conditions and are challenging and expensive to produce. Smaller immunoglobulin fragments, e.g antigen-binding fragment (Fab) and a single-chain variable fragment (scFv), also inherit many of these limitations (Plückthun 2015).

Protein engineering techniques are often employed to improve and develop new non-immunoglobulin-based affinity proteins. This is done for the purpose of minimizing or even replacing the use of antibodies and antibody fragments. Engineered protein scaffolds are attractive alternative to the existing antibody engineering. Scaffolds use different protein frame works that act as a base for amino acid changes with a goal to create new protein variations. What scaffold choice is the best depends on the application and there are already several unique protein scaffolds have been developed and reviewed (Škrlec et al. 2015, Löfblom et al. 2011). What these alternative scaffold -based library technologies need to reach widespread success, is developing robust laboratory tools and optimized techniques.

Currently there are over 20 types of protein scaffolds available. The most common ones are adnectin, affibody, anticalin and DARPin (Luo *et al*. 2022). Monobodies, renamed Adnectin in 2007, are based on the tenth type 3 domain (10FN3) of human fibronectin (Chandler *et al*. 2020). Structural features are similar to variable heavy chains of antibodies exhibiting three exposed loops at one end that are analogous to the antibody variable domain CDRs. The first monobody phage display libraries were generated by randomizing residues within these loops (Koide *et al*. 1998). Over 50 adnectin molecules have found applications in biological or medical research (Luo *et al*. 2022). Affibody scaffold is based on a small isolated and engineered IgG-binding domain denoted as protein Z that was found from *Staphylococcal* protein A (Nilsson *et al*. 1987). Affibody phage display libraries are usually generated by randomizing 13 soluble surface residues

on the first two helixes of the three-helix bundle (Nord *et al*. 1995). The third helix has no binding function but contributes to the structural integrity. A 2-helix affibody only has the binding helixes but compensates the stability loss with disulfide bridges or site mutation offered advantages (Webster *et al*. 2009). Anticalins are scaffolds engineered from lipocalins, which are small extracellular proteins. Four variable surface exposed loops form a ligand binding pocket at the open end of a β-barrel structure. Diversity can be introduced to the design through the sequence and the length of these loops. (Rothe *et al*. 2018.) Selection is often done via phage display, but bacterial surface display or ribosomal display can be used too (Beste *et al*. 1999; Richter *et al*. 2014).

Designed ankyrin repeat protein (DARPin) is one of the most advanced alternative scaffold proteins capable of recognizing targets with specificities and affinities equal to antibodies. They are derived from naturally occurring ankyrin proteins, which are frequent in nature and are more predominantly found in eukaryotes than in prokaryotes (Andrade *et al.* 2001). These proteins are even seen as an adaptive immune system in some jawless invertebrates (Pancer *et al.* 2006). Ankyrin repeat proteins are found mediating important protein–protein interactions intracellularly, extracellularly and in membrane-bound form (Li *et al*. 2006). Darpins are not the first repeat protein that has gone through protein engineering, but they are the most advanced in terms of applications. The very same technologies used to study recombinant antibodies can be harnessed to create new repeat binding proteins. Darpins derived from ankyrin repeat proteins are small for easy manipulation and production with simple functional structure showing high thermodynamic stability against heat or denaturant induced unfolding (Binz *et al.* 2004). DARPins are only 14–17 kDa in size compared to 150 kDa size of human IgG molecules, which makes penetrating tissue and distribution more effective. The lack of any disulfide bonds or prosthetic groups makes them easy to produce in bacteria. Darpins display high thermodynamic stability protecting them from heat or denaturant induced unfolding and keep a low aggregation tendency even at high protein concentrations (Binz *et al.* 2004). These characteristics open wider possibilities in advanced formats or applications than antibodies currently reach. The properties that make DARPins appealing for many applications can also work against them as their size and kinetic properties impact negatively on their suitability as protein therapeutics. Human IgG molecules have half-lives of days to weeks and by further engineering it can be increased to several months (Vaccaro *et al.* 2005). The half-life of DARPins can be extended by conjugating them

with hydrophilic molecules, but such measures would increase the mass and eliminate the advantages DARPins have in terms of small size and low molecular weight, not even mentioning the effect on DARPin molecules stability (Boersma 2018).

Darpins usually consist of two to five stacked repeats composed of 33 amino acid residues (Li *et al.* 2006), though even 29 consecutive repeats are possible (Walker *et al.* 2000), each forming long helix-turn-helix motifs followed by a β-hairpin loop. Most interactions between DARPin and a target protein involve the β-loop. Due to intrinsic instability, a single ankyrin repeat cannot achieve folded structure and so at least two repeats are required to form a hydrophobic core (Zhang *et al.* 2000). As the length of the DARPin increases, so does its thermodynamic stability. Having more than three internal repeats makes DARPin very stable against temperature or chemically induced denaturation. (Wetzel et al. 2008.) A hydrophobic interface keeps these repeats modules together. This interface is supported by the first (N-capping repeat or N-cap) and the last repeat (C-capping repeat or C-cap), which present a hydrophilic outside surface exposed to the solvent and protect the hydrophobic core from the hydrophilic environment (Interlandi *et al.* 2008). These adjacent repeats establish target interaction interface giving way to high-affinity interactions though DARPins prefer to bind to structured domains over unstructured peptides (Strubel *et al.* 2022). This right-handed solenoid structure forms a large solvent facing surface and a continuous hydrophobic core (Sedgwick *et al.* 1999). Singular repeats can be substituted, removed, or inserted without compromising the domains tertiary structure. The general structure of the protein coils into a superhelix, creating a groove-like binding surface (Parmeggiani *et al.* 2017). DARPins have a larger surface area to volume ratio than globular proteins due to this elongated structure, an advantage in protein-protein interactions (Grove *et al.* 2008). The linear manner, in which the repeats assemble, produces a simple scaffold where the most common interactions are hydrophobic short-range interactions within repeats or between the adjacent repeats (Perez-Riba *et al.* 2018). In general, there is no interaction between residues of non-adjacent repeats. Repeat proteins lack many stabilizing long-range interactions seen in globular proteins. These properties make it easier to study DARPin protein folding, stability, or effects of post translational modifications on them (Kane *et al.* 2021).

When building DARPin libraries, consensus strategy is often used to identify important residues, such as residues that are necessary for protein stability and should be kept as is,

or residues that might interact with the target and can be safely randomized (Forrer *et al.* 2004). Library randomization schemes often avoid residues Cys, Pro and Gly (Virnekäs *et al.* 1994). Proline and glycine are avoided for structural hostility and cysteine because it forms disulphide bonds. Structurally important framework positions are often left fixed in the library, while solvent facing residues that could interact with the target are made variable. Available 3D structure data indicates that majority of randomized positions in the library design contact the target in DARPin target complexes (Gilbreth *et al.* 2012). Ribosome display and phage display are the two most common display and selection methods for DARPin libraries. It is possible to display a library consisting of billions of DARPin molecules on a phage or ribosome surface and this allows for separation of DARPin molecules that bind to a molecule of interest.

Stumpp *et al*. reviewed the role of DARPins in drug discovery and drug development as antibody alternatives (Stumpp *et al.* 2007). DARPins can be developed through many different protein engineering approaches, for example, a combination of library-based strategies with structural biology techniques like crystallization. The repeating modules in repeat proteins share structural and sequence similarities making it possible to design an optimal module while still retaining the characteristics of the family. This theory has been applied to several repeat protein scaffolds, like alpha-helicoidal repeat proteins (Valerio-Lepiniec *et al.* 2015), armadillo repeat proteins (Reichen *et al.* 2014) and tetratricopeptide repeats (Speltz *et al.* 2015). However, the DARPins are the most advanced repeat protein scaffold in biochemical and biomedical applications (Plückthun 2015) and therefore are often the main focus of technological development and review.

1.2 Brief history of DARPins

The discovery of ankyrin repeats dates to 1987, when they were found in *Drosophila* Notch protein and yeast Cdc10 (Breeden *et al.* 1987). The 33 amino acid repeat structure was identified and later found in other molecules, eventually being named after the human cytoskeletal protein ankyrin (Lux *et al.* 1990). Ankyrin repeat proteins form one of the largest repeat protein families. How many repeats a single DARPin protein has differs a lot, but generally these proteins consist of four to six repeats. (Galpern *et al.* 2020.)

DARPins were first designed and developed at the University of Zurich in the laboratory of Andreas Plückthun two decades ago (Binz et al. 2003). The goal was to implement a

novel strategy for harnessing the modular nature of repeat proteins for the generation of DARPin libraries. They designed an ankyrin repeat module composed of fixed framework positions to ensure structural integrity and randomized potential interaction positions. Their theory was that the repetitive architecture of DARPins permits adaptation of their size and thus a strategy was developed that would exploit this modular architecture for the generation of combinatorial libraries of DARPins with variable and modular binding surface to a target protein. The strategy consisted of designing a self-compatible repeat module for a given repeat type. Repeat domains were created by sealing a continuous hydrophobic core composed of two, three or four designed modules between N and C-terminal capping repeats. The assumption was that this would generate repeat protein libraries of varying lengths with very large and highly diversified interaction surfaces. All proteins tested in the study had good biophysical properties: expression in soluble form, large yields, easy purification, monomeric and heat resistant. These properties are attributed to the ankyrin repeat framework, making it a stable and modular protein–protein interaction motif. Confirming it is possible to build modular and stable proteins with randomized surfaces, it was a great starting point for novel binding molecules. (Binz *et al.* 2003)

Consensus design made it possible for further protein engineering. Consensus DARPins went through structural and biochemical analysis, which displayed good expression in E. coli, correct folding and improved thermodynamical stability over their natural counterparts (Kohl *et al.* 2003). Combinatorial libraries were generated based on the consensus design and along with developing appropriate in vitro selection methods to isolate molecules with good affinity against the ligand of choice (Binz *et al.* 2003). Composed of structurally important and non-conserved residues, the consensus sequence has six positions for randomization with any amino acid except cysteine, glycine and proline. N- and C-terminal capping ankyrin repeats shield the two or three repeats long library module from the solvent. The library variability was increased further by randomizing N- and C-terminal capping residues after it was realized that both N- and C-terminal capping repeats have a role in target binding (Interlandi *et al.* 2008).

The first therapeutic DARPin was MP0112 for the treatment of wet age-related macular degeneration (Souied *et al.* 2014). The expertise gained from MP0112 was later used at generating a multi-domain dual inhibitor MP0250 for the treatment of cancer, the first multi-domain DARPin tested in patients (Fiedler *et al.* 2017, Binz *et al.* 2017). Ensovibep

was developed for the treatment of COVID-19 and is the first antiviral DARPin overgone clinical trials (Rothenberger *et al.* 2022). Ensovibep is made of single chain of five DARPin binding domains, two domains being human serum albumin (HSA) binding to extent in vivo half-life and three domains that bind the receptor-binding domain (RBD) of the SARS-CoV-2 spike trimer. The addition of HSA binding domains to DARPins improves pharmacokinetic properties and serum half-life as a smaller DARPin molecule would be rapidly cleared by kidney filtration (Steiner *et al.* 2017). All three RBD binding DARPin domains target a common epitope with high affinity and specificity while having different paratopes and through this design ensovibep is resistant to mutations in the domain binding site as the protein can also bind with high avidity (Rothenberger *et al.* 2022). Linking all these different DARPin domains has no negative effect on folding, expression or biological function (Steiner *et al.* 2017, Binz *et al.* 2017).

1.3 Innovations made to the DARPin design

In original library designs, both hydrophilic capping repeats N-cap and C-cap were taken straight from a natural ankyrin repeat protein (Binz *et al.* 2003), guanine-adenine-binding protein, and this limited the stability of the initial DARPin design. Molecular dynamics experiments performed on DARPins revealed that in most high-temperature simulations the C-cap was the first part to denature (Interlandi et al. 2008). C-cap was then redesigned to be more resistant to unfolding caused by denaturants and thermal conditions. This redesign more closely resembled the consensus ankyrin repeats. This redesign was analyzed with nuclear magnetic resonance spectroscopy (Wetzel *et al.* 2010) and crystallography (Kramer *et al.* 2010), which revealed that this stabilizing effect was due to better packing. Introduction of several point mutations to the redesigned C-cap has increased the stability of DARPin further. By introducing five additional mutations to DARPin with three full-consensus internal repeats, with mutations located in the interface to the preceding repeat, the protein showed a rigid-body movement of the C-cap towards the internal repeat (Kramer et al. 2010). Compared to the original C-cap, this version had an increased buried surface area, a superior surface complementarity and improved stability in equilibrium unfolding. The addition of three mutations to C-cap introducing suitably spaced charged residues failed to increase stability as there was no formation of salt bridges. These modifications to the C-cap design allow more

randomization in the C-cap, giving DARPin a larger interaction surface (Interlandi *et al.* 2008).

The architectural concept of DARPins has been expanded with the creation of LoopDARPins (Schilling *et al.* 2014). DARPins bind to target through rigid-body interactions with DARPin backbone experiencing no conformational changes to it. In theory, the concave shape, rigidity, and incompletely randomized binding surface of DARPins limits the amount of possible target epitopes. There might exist epitopes that require more flexibility in binding. A 19-amino acid consensus designed elongated loop, inspired by an antigen-binding CDR-H3 loop found in antibodies, was introduced into the DARPin scaffold at the central β-turn in the second internal repeat. This created a continuous convex paratope common in antibodies. The top of the loop has conformational flexibility, while the loop as a whole contains ten variable positions. A continuous interaction surface is formed as this inserted loop is surrounded by randomized β-turns of neighboring repeats. To make sure the inserted loop did not clash with the neighboring repeats, the stem was kept constant. In creation of LoopDARPin library, the interaction surface was extended further by randomizing both N- and C-capping repeats. Biophysical characterization showed that the introduced loop did not decrease the stability of the scaffold. Binders with affinity as low as 30 pM were isolated from ribosome library after only one round of display with no decrease in scaffold stability.

Position 17 of the N-terminal capping repeat was identified as important component of overall protein thermostability. Asp17 was replaced by either Leu, Val, Ile, Met, Ala, or Thr, which lead to 8 °C to 10 °C increase in the melting temperature of a DARPin domain. Asp17Leu mutation was transferred to other DARPin domains, and lead to improved thermostability in all cases. The modified DARPin domains experienced reduced flexibility and increased stability on account of reduced electrostatic repulsion and improved van-der-Waals packing. (Schilling *et al.* 2022.) This Asp17Leu mutation might be behind SARS-CoV-2 entry inhibitor DARPins, ensovibep, very high melting temperature of >90 °C as it is present in the N-terminal caps of three of the five DARPin domains (Walser *et al.* 2020).

The randomized positions in DARPin library scheme can be converted into consensus residues to create full consensus DARPins (Wetzel *et al.* 2008). The choice to fix positions is based on structural considerations and sequence statistics. The end result

should yield a very high stability protein that can endure many changes to its function, even when those changes are simultaneously detrimental to its stability, making an interesting starting point for testing experimental and theoretical views (Binz *et al.* 2006). These full consensus DARPins have been shown to express well in *E. coli* with their stability increasing as more repeats are added. At more than three internal repeats, the protein becomes even more resistant to boiling or guanine hydrochloride induced denaturation (Wetzel *et al.* 2008). These full-consensus DARPins have higher thermodynamic stability than members of consensus DARPin libraries or naturally occurring ankyrin repeat proteins. Structural analysis suggests that the increase thermostability in the full-consensus DARPin is primarily caused by salt bridge network on the surface. This also implies that surface-exposed electrostatic interactions and regular charge networks have a stabilizing effect. (Merz *et al.* 2008.)

The unique architecture of DARPins is well suited for versatile rigid fusions. Both the N-terminal and C-terminal ends of DARPin form α-helices, which makes it possible to fuse with any other helical scaffold by either extending or overlapping these helices (Batyuk *et al.* 2016). The DARPin termini are positioned laterally to the binding site and so a variety of architectures can be fused while still being compatible with target binding. The advantages of this fusion method are in the control over the directionality of the fused scaffolds, e.g. geometry of epitopes, and in rigidity. By consecutively joining two helices in a long "flagpole" helix gives a non-rigid fusion. This strategy also be used to create DARPin-DARPin rigid fusions with DARPins that are helical on either side (Wu et al. 2017).

1.4 Darpin library selection methods

Synthetic DARPin libraries are compatible with most display or selection methods, but the two most used techniques are ribosome display and phage display. For DARPin selections, ribosome display is the most used method. Ribosome display is in vitro method that physically links genotype (mRNA) and phenotype (encoded protein) through a stalled ribosome. This protein-ribosome-mRNA complex then binds to an immobilized target in a selection step (Hanes *et al.* 1997). The linkage between phenotype and genotype can reliably be maintained only in low temperatures, in non-denaturing conditions and in the absence of RNAses (Zahnd *et al.* 2007a). The method was first used

to improve scFVs (Hanes *et al.* 2000), but has expanded to other scaffolds, like DARPins. Ribosome display has been used to select high-affinity DARPin binders against model ligands such as maltose binding protein (Binz *et al.* 2004), but also against more clinically relevant targets like breast tumor related protein HER2 (Zahnd *et al.* 2007b). Ribosome display yields large sized (10^{12}-10^{14}) and highly diverse libraries to select and evolve protein binders from against chosen target (Leemhuis *et al.* 2005). Ribosome display incorporates PCR into its procedure allowing incorporation of further randomization steps using different error-prone strategies (Dreier *et al.* 2011a). Darpin libraries selected with ribosome display seem to enrich faster than comparable scFv libraries as DARPins fold well in vitro (Dreier *et al.* 2011b). This is attributed to DARPins folding in cell free translation and to their rigid fold, which is resistant to mutation induced instability. Utilizing ribosome display, scientist have managed to enrich DARPin binders with affinities to dissociation constants (Kd) in the picomolar range (Zahnd *et al.* 2006). This was done through efficient off-rate selection, which can be optimized with computational analysis (Zahnd *et al.* 2010).

SNAP display shares similarities with ribosomal display as it is an in vitro selection strategy that replaces reliance on mRNA with more stable DNA (Houlihan *et al.* 2013). Genotype and phenotype are linked via a SNAP-tag in an emulsion microdroplet, which is based on the O^6-alkylguanine-DNA alkyltransferase. SNAP-tag covalently binds DNA templates labeled with benzylguanin substrate. Improving the performance of DARPin display and optimizing conditions for SNAP display led to HER2-binding DARPins being selected at 10^7-fold enrichment rate (Houlihan *et al.* 2014). Even though high affinities have been achieved via SNAP display, the method seems to carry a trade-off between affinity and thermal stability (Houlihan *et al.* 2015).

Filamentous phage display establishes, just like ribosome display, a physical linkage between a polypeptide and the encoding genetic information. Encoding DNA is enclosed in a phage capsid with the protein being displayed on the surface. Mutations introduced to the encoding DNA cause changes in the displayed protein, which is the foundation for generating a library of related variants (Smith 1985). The technique has seen extensive use for selection of peptides or antibodies with desired characteristics (Winter *et al*. 1994), though DARPins have been shown to suffer from poor display when using standard phage display vectors (Steiner *et al*. 2006). In filamentous phage display, the DARPin is fused to minor coat protein p3. The *E. coli* Sec system first produces a membrane bound

intermediate. Majority of this intermediate is secreted to the periplasm while the p3 stays attached to the inner membrane by its C-terminal helix. Afterwards, the coat of the extruding phage picks up the whole fusion protein. (Plückthun 2015.) Compared to ribosome display, the linkage in filamentous phage display is very robust, owning to the fact that the displayed protein is physically linked to the surface of a very stable phage particle encapsulating the corresponding genotype (Steiner *et al.* 2008). Advantages phage display offers over ribosome display are in greater variety of pH or temperature conditions and possible biopanning on whole cells (Steiner *et al.* 2008; Siva *et al*. 2008). The big limitation filamentous phage display has is the bacterial transformation required to produce the protein displaying phage particles as this bacterial transformation limits the library size to below 10^{11} members (Steiner *et al.* 2008). Another issue is the fact that the display requires the protein go through successful translocation to the periplasm and folding (Steiner *et al*. 2006).

A successful phage selection experiment has multiple parameters. General rule of thumb is that larger libraries have higher chance to isolate diverse set of binders against a target protein and will yield higher affinity binders (Ling 2003). Functional library size is the number of different and correct molecules that can be selected. Often the functional library size is but a fraction of transformed library size. Different factors affect functional diversity: the selection systems compatibility with the displayed protein, the library design and the size and quality of the transformed library. Darpins, being a very stable scaffold, should be highly compatible with phage display. They express well, highly soluble, stable thermodynamics and fast folding.

The poor display level of DARPins, when using phage display, is attributed to fast cytoplasmic folding (Wetzel *et al.* 2008). This folding is thought to be too fast for DARPins to be transported across the cell membrane via post-translational Sec system, which is the standard way for *E. coli* protein secretion. This has given rise to the belief that DARPins must be secreted via the cotranslational signal recognition particle (SRP)-dependent system (Bibi 2011) to have high display rates. Enrichment rates for phage display selection of DARPins utilizing SRP-dependent signal sequence were as fast as scFv fragments using conventional Sec mechanism while also improving the display levels of DARPins up to 700-fold (Steiner et al. 2006). DsbA is the current standard signal sequence for filamentous phage display of DARPins. DsbA leader peptide utilizes SRP pathway to facilitate cotranslational translocation of proteins. By utilizing the SRP-

dependent system in constructing DARPin libraries, scientists have managed to isolate DARPin binders with subnanomolar dissociation constants without any affinity maturation (Steiner *et al.* 2008). However, this study and the research done by Kulmala at the University of Turku have raised doubts about the validity of these claims about the effectiveness of Sec and SRP-dependent systems related to the phage display of DARPins. Kulmala *et al.* screened two post-translational PelB signal sequence libraries consisting of leader peptide variants to identify mutants with improved display of DARPins on phage (Kulmala *et al.* 2022). The best PelB variant had synonymous mutations in the n-region and hydrophobic substitutions in the h-region that increased the display efficiency of a DARPin library 12-fold compared to DsbA. This variant had increased cotranslational translocation capabilities compared to PelB wild type, but this could not entirely explain the overall display efficiency improvement over DsbA suggesting that other factors must be contributing to it. For phage display, the post-translational pathway has a greater transportation capacity than the SRP-pathway. SRP-pathway also suffers from overloading liability, which makes it potentially toxic to the host cell. Combinatorial DARPin libraries can be selected by both phage display and ribosome display in parallel and obtain different binders by both methods (Stefan *et al.* 2011).

E. coli does have a third transport system in the twin arginine transport (Tat) system, but all attempts to utilize this route for functional display of p3 fusions have failed (Nangola et al. 2010). It is theorized that the full-length p3 protein is incompatible with the Tat system. There has been some success at Tat-mediated phage display with a truncated version of p3 (Speck *et al.* 2011).

Successful DARPin selection has been achieved utilizing yeast surface display for DARPin domain that recognizes malachite green and functions as a fluorogen-activated protein (FAP) (Schütz *et al.* 2016). Darpin display rates compare well to proteins with high display level on yeast and this effect is probably caused by DARPin being compatible with the yeast quality-control system located at the endoplasmic reticulum. Yeast display is limited to smaller libraries than phage or ribosomal display but has been used to effectively select clones by fluorescence-activated cell sorting (FACS) (Schütz *et al.* 2016). Introduction of FACS to the selection process does give yeast display great potential as a complement to the other display methods (Boersma 2018).

The last amongst the display methods mentioned here is protein fragment complementation. Method has been successfully used in combination with ribosome display to select specific MAP kinase-binders (Amstutz *et al.* 2006) with affinities in the low nanomolar range. In protein fragment complementation system, two halves of dihydrofolate reductase are brought together as a function of target recognition and the selection occurs in cytoplasm. Proteins with potential interaction are each fused to one of the complementary halves of the enzyme dihydrofolate reductase and co-expressed in E. coli. Interaction between these proteins reconstitutes the enzyme and in any bacterial cell where this occurs can grow on selective media.

1.5 Applications of DARPins

Darpins have potential in various biological applications, such as imaging, as fluorogen-activating proteins (FAP). Most FAPs reported to date are scFvs that rely on specific dye interactions, but ScFvs cannot fold accurately without intramolecular disulfide bonds and have lower stability to thermal stress than full antibodies (Wörn et al. 2001). FAPs based on DARPins make up for the scFv shortcomings with their lack of any disulfide bonds or prosthetic and high stability. Schütz and his team isolated DARPins that act as FAPs by activating a fluorophore called malachite green through specific interaction. By adding malachite green to select DARPins, they were able to detect significantly enhanced fluorescence signals. Upon further study, the interaction between malachite green and the select DARPin caused a formation of a homodimer. One select binder was demonstrated to function as a marker for labeling both intracellular and extracellular proteins. (Schütz *et al.* 2016; Könning *et al*. 2018).

The feature of DARPins that makes them attractive option for biosensors is being able to be expressed in a functional form inside the cell. Utilizing ribosome display, DARPins specifically binding to mitogen-activated protein kinase ERK2 in both of its non-phosphorylated and doubly phosphorylated form (Kummer *et al.* 2012). These DARPins did not bind to other kinases tested. The specificity inside the cells was verified by bioluminescence resonance energy transfer. The DARPins bind to the same region of the kinase, but at different activation-states as each DARPin recognizes the conformational change within the activation loop and an adjacent area that occurs upon kinase activation.

Intracellular signaling could be studied in real space and in real time with this kind of activation-state-specific sensors and kinase-specific inhibitors.

Rigidity of DARPins eases crystallization with their respective targets (Bukowska *et al.* 2013) with this facile crystal formation having potential in studying the molecular interactions in atomic detail. X-ray crystallography produces major structural information, a process that could be made easier by engineering DARPins just for this purpose. Cocrystallization between a DARPin and protein that otherwise does not crystallize depends on whether DARPin can dominate crystal packing. To optimize DARPins, rigid domain-domain fusions of the DARPins to larger protein have been generated (Batyuk *et al.* 2016). Rigid multidomain DARPins could address the crystallographic phase problem in molecular replacement by using such rigid DARPin fusions.

Darpins make for interesting binders for diagnostic applications as the advantages in robustness, and the fact that the molecular format can be freely chosen, give a lot of freedom when designing fusion and conjugate proteins (Plückthun 2015). Darpins have a chance in replacing antibodies, when developing new diagnostic detection systems. A DARPin specific to human epidermal growth factor receptor 2 (HER2) was compared to a US FDA-approved rabbit monoclonal antibody to prove its merit in quantitative immunohistochemistry (Theurillat *et al.* 2010). Testing was conducted to investigate DARPins ability to identify the amplification status of the HER2. The comparison was done on breast cancer tissue microarrays and HER2 amplification status was measured by fluorescent in situ hybridization. Darpin and antibody correlated strongly with each other in terms of amplification status and HER2 expression measured with the DARPin having the edge over antibody. This gives solid evidence for the use of DARPins for diagnostic pathology. The affinity of the DARPin was found crucial, but already being a picomolar binder, it was more advantageous to keep it monovalent form rather than make it multivalent construct to increase avidity.

Viral retargeting strategies could benefit from new approaches made possible by using DARPins. Bispecific adapters containing two fused modules, both consisting of DARPins, were developed (Dreier *et al.* 2013). These adapters had extremely high affinity to virus and could stay bound for more than 10 days. The adapter is composed of two fused DARPin modules with one module binding to the fiber knob of adenovirus serotype 5 and the other binding to a cell surface receptor. This interaction blocks the

natural receptor-binding site of the virus and mediates interaction with a surface receptor of choice, e.g., various possible tumor markers. These adapters can be produced in E. coli and when coupled with adenoviruses converted to new receptor specificities, allows adenoviruses with different payloads to be retargeted to many different cell types of choice. This overcomes some of the main challenges facing viral delivery: to specifically target the virus to the cells of interest and deliver a payload best suited for the purpose (Plückthun 2015).

Proteins participating in cellular regulation are often multi domain, making a bispecific binder that targets different domains of the same protein a very attractive idea (Sha *et al.* 2017). Out of all biparatopic non-immunoglobulin scaffold binders designed so far, DARPins might be the most explored group. The potential of DARPins for this application was first shown by a dimeric DARPin composed of two select DARPins binding to two different FcεRIα epitopes. These epitopes are involved in the IgE interaction, and the purpose was to inhibit IgE-FcεRI complex formation (Eggel *et al.* 2014). These two DARPins were fused together via Gly-Ser linker in two different configurations with one configuration showing enhanced affinity compared to monovalent DARPins. Fusion of two DARPins has also been utilized in targeting two different epidermal growth factor receptor (EGFR) epitopes in an attempt to inhibit the growth of cells expressing EGFR (Boersma *et al.* 2011). Two different fusion strategies were used to analyze two different biparatopic formats. The first format fused two DARPins with a flexible Gly-Ser linker and the second format was simultaneously biparatopic and bivalent design utilizing a more rigid, non-covalent, leucine zipper motif to link together two copies of each DARPin. The construct with Gly-Ser linker was similar to monovalent control proteins in efficacy, but in contrast the leucine zipper construct was more effective at inhibiting proliferation and cell viability than the combination of the two monovalent control DARPins. In both above studies, the orientation of the DARPin modules was important for simultaneous binding activity. A study indicates that more rigid Pro-Thr linkers might also have an advantage over the more classical Gly-Ser linkers (Steiner *et al.* 2017).

DARPin molecules have many good qualities to capitalize on: high stability, solubility, and aggregation resistance. These empower scientists to create a wide range of conjugates and fusions and are the reason for all the above advancements.

1.6 Aims of the study

Goal of the study was to first analyze signal sequence candidates in a library context, choose the best performing signal sequence variant and construct a combinatorial DARPin library utilizing it. This DARPin library should fulfill all the parameters of a successful phage selection experiment. The first task was to improve phage display technique by finding advantageous PelB signal peptide variants. Previous studies analyzing the display efficiency of PelB signal sequence variants were conducted with just anti-GFP DARPin (Kulmala 2020). The display efficiency of PelB signal sequence variants on phages was evaluated by linking them to binary library DARPins to confirm that the variants identified in Kulmalas research could improve the display efficiency of DARPins other than the anti-GFP one. The combinatorial DARPin library was then constructed on the best performing PelB signal variant and the success of library construction was evaluated by panning it against protein N from SARS-COV-2

2. Materials and methods

2.1 Creation of PelB signal sequence variants with two synonymous mutations

PelB signal sequence variants with one synonymous codon alteration, in either in the n-region and another in the hydrophobic region of PelB signal sequence, were from Dr. Antti Kulmala (Kulmala 2020). To analyze the effect of two synonymous codon alterations, PelB signal sequence mutants DN5 and DN10 were combined with DH4, creating DN5-DH4 and DN10-DH4 variants. To create pEB32x vector backbones with double mutant PelB signal sequences, pEB32x-DARPin PelB DN5 and DN10 vectors were amplified with PCR, both using common forward primer TH361 (Table 1.) and mutant-specific reverse primer: TH359 and TH360 (for DN5–DH4 and DN10–DH4, respectively). The PCR amplification was performed using Phusion High-Fidelity DNA Polymerase (Thermo Scientific, Waltham, USA) 1 µl per 100 µl PCR mix, 5x Phusion HF Buffer, 0.8 mM dNTPs, 0.5 µM primers and 3-5 ng template DNA. The cycle protocol was carried out on a Bio-Rad PCR machine: 30 sec 98°C (initial denaturation); 35 cycles of 10 sec 98°C (denaturation), 15 sec 66°C (annealing) and 90 sec 72°C (extension); a final extension step 5 min 72°C; cooling to 4°C. PCR samples were then loaded to agarose gel and extracted from an agarose gel with GeneJet Gel Extraction kit (Thermo Scientific).

To ligate the double mutant vector fragments to TetR cassette (stuffer fragment of 2101 bp), the double mutant vector backbones were digested by using 5 U SfiI restriction enzyme (Thermo Scientific) for all gel extracted DNA in 2 h reaction at 50 °C. SfiI digested TetR cassette gene was ligated in 1:3 vector/insert molar ratio by using T4 DNA Ligase (Thermo Scientific). The ligation reactions were incubated for o/n at 16 °C. No heat inactivation as the ATP needed for ligation is consumed during the first few hours. To determine the if cloning was successful, 2 µl of both ligations were mixed with 20 µl *E. coli* XL-1 Blue cells transformed into the cells with Bio-Rad Genepulser (Bio-Rad, Hercules, USA) with settings 1.25 kV, 25 µF, 200 Ω in two separate electroporations. After electroporation, the cells were recovered in 1 ml of SOC medium at 37 °C for 1 h with 250 rpm shaking. 100 µl of the recovered cells were plated on LA plates (0.5% w/v glucose, 25 µg/ml chloramphenicol, 10 µg/ml tetracycline). Plates were incubated at 37 °C, o/n. Four colonies from both plates were inoculated in 20 ml SB culture medium (10 µg/ml tetracycline, 25 µg/ml chloramphenicol) and incubated o/n at 30 °C, 300 rpm.

Plasmid DNA was extracted from o/n cultures with GeneJET DNA Miniprep Kit (Thermo Scientific) and DNA was sent to be sequenced by Macrogen (Seoul, South Korea).

pEB32x-DARPin vectors for parental PelB (wildtype), DN5, DH4 and DN10, as well as pEB32x-Tet double mutant vectors for DN5–DH4 and DN10–DH4, were digested with 5 U SfiI for 1-2 µg of vector DNA, gel extracted and then ligated in 1:3 vector/insert molar ratio with a combinatorial DARPin library SfiI DNA insert containing serine/tyrosine codon variation (TMY) at specified codon positions. This DARPin binary library was purchased as ready linear DNA block from Eurofins Genomics (Ebersberg, Germany) and was amplified with PCR before digesting with SfiI. The primers used in library amplification added SfiI restriction enzyme recognition sites to the sequence. The ligated library DNA was transformed to *E. coli* XL-1 Blue cells with electroporation done like above, except the recovered cells were diluted 10^{-1} for platings. Each transformation yielded >10 000 cfu transformants. Every PelB mutant library had vector control ligation (no insert) prepared parallel, which were plated alongside the libraries.

DsbA-BiliBST library was cloned using Gibson assembly. A template containing DsbA signal sequence, pEB32x-DsbA-DARPin anti-GFP, was amplified with primers TH363 and TH364. The DARPin binary library was amplified with primers TH362 and JLe01as using pEB32x-DARPin library (wildtype PelB) as the template. After amplification, PCR samples were loaded to agarose gel and extracted. Gibson reaction protocol is pretty much identical to the original (Gibson *et al.* 2009), but with T5 exonuclease prediluted 1:10 in Cutsmart buffer. Gibson Assembly Master Mix was thawed at room temperature and 15 µl of it was mixed with 50 ng of both vector and library insert. The reaction was incubated at 50°C in a thermocycler for 1 hour. 2 µl of the reaction was transformed to 20 µl of *E. coli* XL-1 Blue with electroporation yielding 160 000 transformants. Extracted the resulting pEB32x-DsbA-DARPin DNA from the colonies. Another Gibson assembly was performed with 80 ng of vector and 200 ng of library insert mixed with 30 µl of master mix. This reaction solution was purified with DNA Clean & Concentrator kit (Zymo research) and 6 µl of purified DNA in elution buffer was transformed to 50 µl of *E. coli* XL-1 Blue cells yielding 330 000 transformants followed by extracting the DNA from colonies.

The Gibson assembly suffered from high background and multiple attempts were made to lower it. In the end, creating the DsbA-BiLiBST fragment with PCR amplification using primers W0375 and PAKrev with the smaller pEB32x-DsbA-DARPin library as

template proved successful. The PCR amplification was performed using Phusion High-Fidelity DNA Polymerase (Thermo Scientific) 1 µl per 100 µl PCR mix, 5x Phusion HF Buffer, 0.8 mM dNTPs, 0.5 µM primers and 200 ng template DNA. The cycle protocol: 30 sec 98°C; 35 cycles of 10 sec 98°C, 15 sec 60°C and 30 sec 72°C; a final extension step 5 min 72°C; cooling to 4°C. The PCR reaction was purified with PCR purification kit (Thermofisher). 2 µg of purified DNA was digested with Xbal and Hindlll restriction enzymes in o/n reaction at 37 °C followed by gel extraction of the correct size fragment covering DsbA-DARPin-p3-CT gene. The pEB32x-DsbA-DARPin library insert was ligated in 1:3 vector/insert molar ratios to Xbal and Hindlll digested pEB32x vector fragment in 1 h reaction at RT and inactivated for 10 min at 80 °C. 1 µl of ligated library was electroporated to 10 µl of *E. coli* XL-1 Blue cells yielding >10 000 cfu transformants with 3% vector background colonies. The library cloning strategy was confirmed by sequencing ten DN10-DH4 (same insert in all PelB libraries) and ten DsbA clones.

Table 1. Primer sequences

Name	DNA sequence 5'-3'
TH347	CTCGGCCCCCGTGGCCGACGCTGC
TH348	GCGGCCCAGCCGGCCATGGCGG
TH359	CATGGCCGGCTGGGCCGCAAGAAGAAGTAGACCAGCTGCTGCGGTGGGGAGGAGATATTTC
TH360	CATGGCCGGCTGGGCCGCAAGAAGAAGTAGACCAGCTGCTGCCGTCGGTAGAAGGTACTTC
TH361	TCCTGCAGAAACTGGGAGCGG
TH362	TAGCGTTTAGCGCATCGGCGGATCTGGGTAAAAAACTGCTG
TH363	TGAAGGTCGTGGCGGTTCTGGTTC
TH364	ATCCGCCGATGCGCTAAACGC
JLe01as	ACCAGAACCGCCACGACCTTCAA
WO375	TCACACAGGAAACAGCTATGAC
PAKrev	CGCCATTTTTCACTTCACAG

2.2 Construction of PelB variant and DsbA DARPin display libraries

Five batches of phage stocks were produced from the PelB variants and DsbA libraries using the same method. For the first batch, cells were scraped off from the plates with 2

ml of SB medium (2 ml/plate) and diluted to OD (600 nm) 0.1 in 20 ml SB medium (1% w/v glucose, tetracycline, chloramphenicol) and cells were cultured at 37 °C with 300 rpm shaking to OD (600 nm) 0.3-0.6. Cells were infected with 10 µl VCS M13 helper phage ($2x10^{11}$ cfu/ml ($2x10^{14}$ pfu/ml), Stratagene) and incubated at 37 °C for 30 min with no shaking. Cells were cooled on ice and centrifuged at 4000 rpm for 10 min at 4 °C. Supernatant was discarded and cells were resuspended in glucose free SB medium (tetracycline, chloramphenicol, 30 µg/ml kanamycin, 5 mM $MgCl_2$). Incubated cultures at 30 °C o/n with 300 rpm shaking. For the rest of the batches, inoculated cells from glycerol preps to 5 ml SB medium (1% w/v glucose, tetracycline, chloramphenicol) and incubated at 30 °C o/n with 300 rpm. 20 ml main cultures were started from these precultures like above.

Phage stocks were prepared by first centrifuging o/n cultures at 12 000 x g for 10 min at 4 °C. Supernatants were collected and 1/5 volume of PEG/NaCl (20% PEG-8000 and 2.5 M NaCl) was added to precipitate the phages. Precipitation reactions were incubated on ice for 1 h and subsequently centrifuged at 10 000 x g for 20 min at 4 °C. Phage pellets were dissolved in 1 ml of TBS buffer (50 mM Tris-HCl, pH 7.5; 150 mM NaCl) and subsequently centrifuged with tabletop centrifuge at 16000 x g for 5 min at 4 °C to remove residual cells. Supernatants were transferred to new tubes and reprecipitated with 1/5 volume of PEG/NaCl. The phage samples were incubated on ice for 5 min and centrifuged at 10 000 x g for 5 min at 4 °C. The supernatant was removed, and the phage pellets were dissolved in 0.5 ml of TSA/BSA buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 0.02% w/v Na-azide, and 1% w/v BSA). Phage stocks were stored at 4 °C for use.

Phage titers for the first three batches of phage stocks were determined by oligonucleotide-directed chelate complementation assay (OCCA) (Lehmusvuori *et al.* 2013). The phage quantification method is based on measurement of phage ssDNA using lanthanide chelate complementation probes. First, phage stock samples were diluted 1/100 in TBS and prepared pEB32x-ScFvP standards diluted to $4x10^{11}$, $2x10^{11}$, $1x10^{11}$, $5x10^{10}$ and $2x10^{10}$ cfu/ml in TBS. Mixed 10 µl of sample with 60 µl ready aliquoted probe master mix (10 nM Eu-oligo + 5 nM antenna-oligo, 25 mM Tris-Cl pH 7.5, 900 mM NaCl, 0.1% Tween-40, 0.05% NaN3, 30 µM DTPA) and heated he sample-probe mix for 1 min at 95 °C and cool to 25 °C in PCR cycler. Applied 60 µl of sample-probe mix on yellow C-12 low fluor Maxi microtiter wells (Kaivogen, Turku, Finland) and incubated strips for 15 min with shaking. After incubation, time-resolved europium signal was

measured with Victor 1420 Multilabel Counter. Calculated linear fit to standard and transformed signal to phage titer with the derived equation. Phage titers for the last two batches of phage stocks were determined by total phage immunoassay described in the next section. This method was also used to redetermine the phage titers for the first three batches as well.

2.3 Analyzing the phage display levels of PelB signal sequence variants

Before analyzing the phage display levels of PelB signal sequence variants, the phage stocks titers were normalized by total phage immunoassay. DsbA-BiLiBST phage stock from the first batch was used as a standard to estimate the number of phages and its OCCA results were assumed to be accurate enough. It was used as the reference for determining the other phage stock titers and this way all phage stocks have the same amount of systematic error. Wells of a 96 well streptavidin plate (Kaivogen) were coated with 100 µl per well of 250 ng/ml biotinylated anti-M13 monoclonal antibody (University of Turku, Turku, Finland) in Assay Buffer (Kaivogen). The plate was incubated was incubated at RT for 30 min with slow shaking. After incubation, the plate was washed twice with Delfia Plate Wash by using Kaivogen wash buffer. After washes, 100 µl per well of phages as triplicate from each library phage production were added diluted to $5x10^6$ cfu/ml in Assay Buffer alongside standard set of $5x10^9$, $5x10^8$, $5x10^7$, $5x10^6$, $5x10^5$, $5x10^4$ and $5x10^3$ cfu/ml. Incubation and wash steps were done as above. Bound phages were detected with 25 ng per 100 µl europium labeled mouse anti-phage antibody. Incubation again as above, but after incubation the plate was washed four times. Subsequently, 200 µl of Enhancement solution was added and the plates were incubated at RT for 10 min with slow shaking followed by measuring the time-resolved europium signal. Calculated linear fit to standard and transformed signal to phage titer with the derived equation. Based on the first phage titer results, phage titers were readjusted and analyzed again with the method above 2-3 times until the phage titers were within 20% of each other.

After normalizing the phages from each library phage production, the capability of each signal sequence to display serine/tyrosine binary library DARPins on the surface of phage particles was determined by anti-DARPin immunoassay. Wells of a 96 well streptavidin plate were coated with 100 µl per well of 250 ng/ml biotinylated 2A11 anti-human Fab

(Hytest Ltd, Turku, Finland) in Assay Buffer. The plate was incubated at RT for 30 min with slow shaking. After incubation, the plate was washed two times. After washing, 100 µl per well of 600 ng/ml anti-DARPin Fab 245E6 (Department of Biochemistry, University of Turku, Turku, Finland) was added to the wells. The incubation and washing steps were done like above. Using the normalized phage titers, the phages from each library phage production, except for DsbA in each batch, were diluted to $5x10^{10}$, $2.5x10^{10}$, $1.25x10^{10}$, $6.25x10^{9}$ and $3.13x10^{9}$ cfu/ml in Assay buffer and added 100 µl per well as triplicate. DsbA had only $5x10^{10}$ cfu/ml sample in triplicate. Incubation and wash steps as above. Bound phages were detected with 25 ng per 100 µl europium labeled mouse anti-phage antibody using the same incubation time as above, but with four washes. Subsequently, 200 µl of Enhancement solution was added and the plates were incubated at RT for 10 min with slow shaking followed by measuring the time-resolved europium signal. Afterwards, total phage immunoassay was done using the phage dilutions from the anti-DARPin immunoassay.

Based on the display level analysis, PelB mutant DN5 was chosen for displaying the Serine/tyrosine binary DARPin library. TetR cassette gene was ligated to PelB-DN5 vector, also named pEB32D9 vector, followed by transformation to *E. coli* XL-1 Blue cells. The correct sequence was confirmed by sequencing two clones.

2.4 Serine/tyrosine binary library DARPin preselection

SfiI digested binary library gene was ligated to Sfil digested pEB32D9 in 1:3 vector/insert molar ratio alongside a vector control reaction. The ligation reactions were incubated for o/n at 16 °C. Ligation reactions were concentrated and cleaned, before transforming ~120 ng of pEB32D9-BiLiBST to 70 µl of fresh *E. coli* XL-1 Blue cells in five separate electroporations. After electroporation, cells were recovered like before and combined all five library transformations. To determine the size of the library, the recovered cells were diluted 10^{-1}-10^{-4} in SB medium, and 100 µl of all dilutions were plated on LA plates. The rest of the cells were plated on 24.5 x 24.5 cm bioassay dish (Sigma-Aldrich, Germany) and incubated at 37 °C, o/n. This process was repeated once to get over 30 million clones, but this time there were seven separate electroporations. The cells were scraped off from the big LA plates with 5 ml of SB medium (5 ml/plate) and combined before diluting to OD (600 nm) 0.1 in 100 ml SB medium (1% w/v glucose, tetracycline, chloramphenicol)

and cells were cultured at 37 °C with 300 rpm shaking to OD (600 nm) 0.3-0.6. Cells were infected with 10 µl VCS M13 helper phage and incubated at 37 °C for 1 h with 300 rpm shaking. After incubation, added kanamycin to final concentration 30 µg/ml and incubated at 30 °C o/n with 300 rpm shaking. Phage stocks were prepared like above.

Preselection of the BiLiBST-library to enrich for open reading frames was done by panning two rounds against a mixture of anti-DARPin Fabs AK005H02, AK014A03, AK017G08 and 245E6 coated on beads as polyclonal mixture. Two different pannings were conducted side by side, one with just TBT-0.05 (TBS, 1% BSA, 0.05% Tween) as panning buffer and another with added 5% w/v fat free milk. Firstly, 20 µl of DynaBeads M-280 Strepavidin (Thermofisher) were washed three times with 0.8 ml TBT-0.05 and resuspended in original volume of TBT-0.05. Then beads were polyclonally coated with four different biotinylated anti-DARPin Fabs (University of Turku), 2.5 µg of each biotinylated Fab and TBT-0.05 to reach 1 ml reaction volume. Incubated for 30 min at RT in rotation. Beads were washed like above and resuspended in original volume. Divided beads for two panning reactions and mixed them with 250 µl of pEB32D9-BiLiBST phages in final reaction volume of 2 ml. Incubated for 1 h at RT in rotation. After incubation, beads were washed three times with TBT-0.05 and once with TBS changing tubes after each wash to lower the background binding. To elute the phages, suspended the beads in 200 µl of trypsin (Sigma-Aldrich) diluted to 60 µg/ml in TBS and incubated for 30 min at RT in rotation. After elution, added soybean trypsin inhibitor (Sigma-Aldrich) to final concentration of 50 µg/ml. Beads were collected with a magnet and the eluate was transferred to a new tube. For infection, 5x volume of fresh *E. coli* XL-1 Blue cells with OD (600 nm) 0.3-0.6 were added on eluate. Infection at 37 °C for 30 min with no shaking. Following controls were also done, infection control with known phage stock and contamination control. The infected cells were diluted 10^{-1}-10^{-3} in SB medium, and 100 µl of all dilutions and control were plated on LA plates. The rest of the cells were plated on three LA plates and incubated o/n at 37 °C. Cells were collected from the plates with 2 ml of SB medium (2 ml/plate). Phage production and phage stock preparation like above. The second panning round was performed like the first one still with two different reactions, but Neutravidin M-270 beads instead of streptavidin. After two rounds of panning, the library preselection strategy was confirmed by sequencing 24 clones from both panning reactions.

2.5 Main Darpin library construction

D2 and D3 motifs were diversified by ordering oligos manufactured by trinucleotide mutagenesis (TRIM) technology (Ella Biotech, Martinsried, Germany) and incorporated into main library using the Ser/Tyr -binary DARPin library as the template (performed by T. Huovinen, Department of Life Technologies, University of Turku). The final assembly PCR product covering the DARPin gene library was digested with SfiI and ligated to pEB32D9 vector. After assembly PCR, the main DARPin library insert DNA was first amplified with PCR using forward primer TH347 and reverse primer TH348. Successful amplification was confirmed by gel electrophoresis. Insert DNA was purified and all of it was SfiI digested before ligated in 1:3 vector/insert molar ratio to pEB32D9 vector. The ligation reactions were incubated for o/n at 16 °C. Ligation reaction was purified with DNA Clean & Concentrator kit (Zymo research) and the ligation success was confirmed by transforming 1 µl of ligation product to 20 µl of E. coli XL-1 Blue cells diluted 1:2 in water. Vector controls were used to calculate background, which was always <10%. For main DARPin library transformation, 1.55 µg of ligated DNA was mixed with 5.25 ml of *E. coli* SS320 cells. Mix was aliquoted to electroporation cuvettes in 350 µl aliquots transformed into the cells with settings 2.5 kV, 25 µF, 200 Ω in 15 separate electroporations. After electroporation, the cells were recovered in 10 ml of SOC medium and pooled together before incubation at 37 °C for 1 h with 200 rpm shaking. In order to determine the size of the library, the recovered cells were diluted 10^{-1}-10^{-5} in SB medium, and 100 µl of the 10^{-3}-10^{-5} dilutions were plated on LA plates twice and incubated o/n at 37 °C. The library size of the first main DARPin library, named DAB B, was $5.7x10^{8}$ cfu. The rest of the cells were suspended in 1 L of SB (chloramphenicol, tetracycline) and incubated at 37 °C for 2 h with 200 rpm shaking before lowering the temperature 26 °C and incubating o/n. The next day, cells were diluted to OD (600 nm) 0.2 in 1 L SB medium (chloramphenicol, tetracycline) and were cultured at 37 °C with 200 rpm shaking to OD (600 nm) 0.6-0.8. Cells were infected with 1 ml of VCS M13 helper phage and incubated at 37 °C for 1 h with 200 rpm shaking. After incubation, added kanamycin to final concentration 30 µg/ml and incubated at 30 °C o/n with 200 rpm shaking. Phage stocks were prepared like usual, but the final pellet suspension was done in 2 ml TSA/BSA. The library transformation was repeated once more, but with 3.1 µg of ligated DNA this time. This second main DARPin library, named DAB C, was $4.3x10^{9}$ cfu.

2.6 The functional validation of the main DARPin library

The functionality of the main DARPin library was validated by panning three rounds and performing phage immunoreactivity assay, which will show if any enrichment has happened as intended. The phages will be panned against biotinylated GST-protein N (Department of Virology, University of Turku, Turku, Finland). For the first panning round, DAB B and C libraries were combined before performing a negative selection in streptavidin wells to remove possible streptavidin binders. Total $5x10^{12}$ cfu of phages were diluted in 2 ml Assay Buffer (Kaivogen) and added phage dilution 200 µl/well. This selection was performed twice, once for the phages used in a panning reaction and one for the phages used in a background reaction. Incubated at RT for 1 h with slow shaking and transferred the phage solution to a tube. Phages were then incubated with 20 µg GST. Incubated in rotation for 1h at RT. 45 µl/reaction of DynaBeads M-280 Strepavidin were washed three times with 0.8 ml TBT-0.05 and resuspended in original volume of Assay Buffer. The beads were then divided between a panning reaction with antigen and a background reaction with no antigen.

The beads for the panning reaction were coated with 4.5 µg of bio-GST-N (University of Turku) and Assay Buffer to reach 1 ml reaction volume. The background reaction beads had only Assay Buffer. Incubated for 30 min at RT in rotation. Beads were washed like above and resuspended in original volume. The bio-GST-N coated and the uncoated beads were mixed with 1 ml of negatively selected and phages with GST. Incubated for 1 h at RT in rotation. After incubation, beads were washed three times with TBT-0.05 and once with TBS changing tubes after each wash. To elute the phages, beads were suspended in 200 µl of trypsin diluted to 60 µg/ml in TBS and incubated for 30 min at RT in rotation. After elution, added soy bean trypsin inhibitor to final concentration of 50 µg/ml. Beads were collected with a magnet and the eluate was transferred to a new tube. For infection, 5x volume of fresh *E. coli* XL-1 Blue cells with OD (600 nm) 0.3-0.6 were added on eluate. Infection at 37 °C for 30 min with no shaking alongside infection and contamination controls. The infected cells were diluted 10^{-1}-10^{-3} in SB medium, and 100 µl of all dilutions and controls were plated on LA plates. The rest of the cells were

centrifuged at 3000 x g for 5 min at 4 °C. Pellet was resuspended in 200 µl and plated to LA plate. Incubated o/n at 37 °C.

The next day, cells were scraped off from the plates with 2 ml of SB medium and diluted to OD (600 nm) 0.1 in 100 ml SB medium (tetracycline, chloramphenicol) and cells were cultured at 37 °C with 300 rpm shaking to OD (600 nm) 0.3-0.6. Cells were infected with 10 µl of VCS M13 helper phage and incubated at 37 °C for 1 h with 300 rpm shaking. After incubation, added kanamycin to final concentration 30 µg/ml and incubated at 30 °C o/n with 300 rpm shaking. Phage stocks were prepared like usual. Before the second panning round, 10 µl/reaction of phages from the first panning round were negatively selected against 5 µl of streptavidin beads. The second panning round was performed in solution. Firstly, mixed 10 µl of negatively selected phages with 6.9 nM of bio-GST-N alongside a background reaction with no antigen. Incubated for 30 min at RT in rotation. After incubation, added 20 µl of washed streptavidin beads to each reaction. Incubated 10 min at RT in rotation after which the beads were washed three times with TBT-0.5 and once with TBS changing tube after each wash. Following this, the protocol was the same as above. The third panning round protocol was the same as the previous round with another negative selection performed before panning, but this time there were two panning reactions, one with 1 nM of bio-GST-N and one with 100 pM of bio-GST-N.

2.7 Analysis of binder enrichment

The enrichment of binders was confirmed with phage immunoreactivity assay. Wells of a 96 well streptavidin plate were coated with 100 µl per well of 250 ng/ml bio-GST-N in Assay Buffer. The plate was incubated at RT for 30 min with slow shaking. After incubation, the plate was washed two times. After washing, unpanned and panned phages were diluted 1/500 in Assay Buffer and added 100 µl per well as quadruplicate to wells coated with antigen and to wells with no antigen. Incubation and wash steps as above. Bound phages were detected with 25 ng per 100 µl europium labeled mouse anti-phage antibody using the same incubation time as above, but with four washes. Subsequently, 200 µl of Enhancement solution was added and the plates were incubated at RT for 10 min with slow shaking followed by measuring the time-resolved europium signal. The enrichment was measured by comparing the signal from wells with antigen to the ones with no antigen.

3. Results

PelB signal sequence variants carrying synonymous codon mutations (Table 2.), with DN denoting mutations in the n-region and DH in the hydrophobic region, were screened for improved display of binary DARPin library compared to DsbA (Kulmala 2020). DN5, DN10 and DH4 were the best variants selected from DARPin N and DARPin H libraries by Kulmala in his research (Figure 1). Five PelB signal sequence mutants were analysed to find the one with the best display levels for binary library. DN5–DH4 and DN10–DH4 variants were created by PCR amplification with mutant-specific reverse primers. The PCR primers used in the amplification insert two Sfil restriction enzyme recognition sites downstream from PelB signal sequence, which will be used in cloning. Sfil restriction enzyme cuts right in the front of PelB signal sequence meaning that, when an insert DNA sequence digested with Sfil is cloned into the vector, it will be under PelB signal sequence.

Table 2. PelB signal sequence variants analyzed in this study

Clone	**M**	**K**	**Y**	**L**	**L**	**P**	**T**	**A**	**A**	**A**	**G**	**L**	**L**	**L**	**L**	**A**	**A**	**Q**	**P**	**A**	**M**	**A**
Parental PelB	ATG	AAA	TAC	CTA	TTG	CCT	ACG	GCA	GCC	GCT	GGA	TTG	TTA	TTA	CTC	GCG	GCC	CAG	CCG	GCC	ATG	GCG
DN5	--G	--G	--C	--T	--A	--G	--G	--A	--C	--T	--A	--G	--A	--A	--C	--G	--C	--G	--G	--C	--G	--G
DN10	--G	--A	--T	--C	--C	--C	--C	--A	--C	--T	--A	--G	--A	--A	--C	--G	--C	--G	--G	--C	--G	--G
DH4	--G	--A	--C	--A	--G	--T	--G	--A	--A	--T	--T	--A	--T	--T	--T	--G	--C	--G	--G	--C	--G	--G
DN5-DH4	--G	--G	--C	--T	--A	--G	--G	--A	--A	--T	--T	--A	--T	--T	--T	--G	--C	--G	--G	--C	--G	--G
DN10-DH4	--G	--A	--T	--C	--C	--C	--C	--A	--A	--T	--T	--A	--T	--T	--T	--G	--C	--G	--G	--C	--G	--G
	M	K	K	I	W	L	A	L	A	G	L	V	L	A	F	S	A	S	A			
DsbA	ATG	AAA	AAG	ATT	TGG	CTG	GCG	CTG	GCT	GGT	TTA	GTT	TTA	GCG	TTT	AGC	GCA	TCG	GCG			

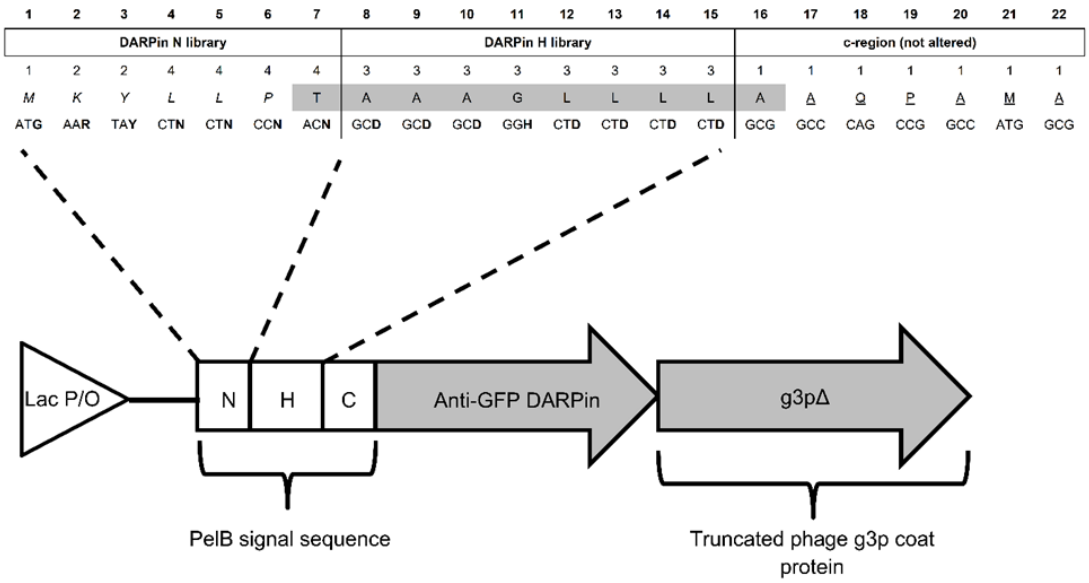

Figure 1. The design of the DARPin N and DARPin H libraries. Vertical bars denote the borders of the libraries. Each library included diversity in the given library region. C-region was not mutated in either library. Grey arrows represent the anti-GFP DARPin attached to the PelB signal sequence, and truncated phage g3p coat protein. White triangle denotes the Lac promoter. Letters N, H and C in the PelB signal sequence denote the n-region, hydrophobic region and c-region, respectively. Positional synonymous codon diversity at each position is shown above the amino acid sequence of the PelB signal sequence. Letters in italics denote the natural n-region, gray shading denotes the natural hydrophobic region and underlined letters denote natural the c-region. H = C, T, A; D = A, G, T; Y = C, T; V = A, G, C; R = A, G (Kulmala *et al.* 2022.)

The DARPin library cannot be cloned to DsbA sequence using SfiI as inserting a recognition site would change the sequence to something else than DsbA. Instead, the DsbA-DARPin library was constructed with Gibson assembly. Gibson assembly was a success with 16 000 transformants, but there was high vector background (30 %). In order to reduce cloning background, i.e., the number colonies with only the cloning vector, the larger pEB32x-DsbA-DARPin library was digested with XbaI, HindIII and BglII restriction enzymes in 2 h reaction at 37 °C. XbaI and HindIII cut DsbA-BiLiBST-gIIIp fragment out and BglII cuts DARPin anti-GFP from the middle if there is any. This method proved unsuccessful as when the digestions were loaded on gels and analyzed under UV light there were only smears instead of proper bands. Afterwards, attempted to create the DsbA-BiLiBST fragment with PCR. The larger pEB32x-DsbA-DARPin library was amplified with primers W0375 and PAKrev, but this also failed. DNA from the larger

pEB32x-DsbA-DARPin library seemed to be of poor quality. The PCR amplification was then tried with the smaller library as template, which ended up succeeding. The library cloning strategy was confirmed by sequencing DN10-DH4 and DsbA clones and, in both library samples, 7/10 clone sequences were according to the library design and 3/10 were frameshift clones.

The binary library contained 23 residue positions with combinatorial variating serine/tyrosine amino acids (Figure 2.). Each PelB signal sequence variant displayed binary library with >10 000 cfu transformants. Every PelB variant library had parallel vector control ligation with no insert DNA, which were and used to calculate < 0.7% vector background for all libraries. The phages produced from these libraries were analyzed by anti-DARPin immunoassay, which measured the specific binding of phages displaying binary DARPin library to anti-DARPin Fab. A batch of phage stocks were produced five times each having a binary library displayed under either DsbA, PelB wild type or one of the five PelB variants. The ratio of PelB variant signal to DsbA signal was used to identify variants with better display of binary library DARPins. To confirm that each sample had about the same number of phages, an anti-phage immunoassay was used to normalize phage titers first. The phage production, immunoassays and analysis were performed in five independent replicates. OCCA was used to determine the phage titers for the first three replicates, but the results from the total phage immunoassay indicate that OCCA lacks accuracy, giving many titers that were significantly lower than according to the immunoassay. These titers were redetermined by total phage immunoassay.

D L G K K L L E A T Z Z G Q D D E V R I L M A N G A D **V** N A Z D Z Z G

GATCTGGGTAAAAAACTGCTGGAAGCAACATMYTMYGGTCAGGATGATGAAGTTCGTATTCTGATGGCAAATGGTGCAGAT**GTT**AACGCCTMYGACTMYTMYGGT

Z T P L H L A A Z Z G H L E I V E V L L K H G A D **V** N A Z D Z Z G Z T

TMYACACCGCTGCATCTGGCAGCATMYTMYGGTCATCTGGAAATTGTTGAAGTTCTGCTGAAACATGGTGCCGAT**GTG**AATGCGTMYGATTMYTMYGGTTMYACC

P L H L A A Z Z G H L E I V E V L L K Y G A D **V** N A Z D Z Z G Z T P L

CCGCTGCACTTAGCAGCATMYTMYGGCCATCTCGAGATCGTGGAAGTTTTACTGAAATATGGAGCTGAT**GTT**AATGCCTMYGATTMYTMYGGTTMYACGCCTTTA

H L T A Z Z G H L E I V E V L L K H G A D **V** N A Q D Z Z G Z T P F D L

CATCTGACCGCATMYTMYGGCCACCTGGAAATAGTAGAGGTGCTGCTTAAACACGGTGCGGAT**GTA**AATGCACAGGATTMYTMYGGCTMYACCCCATTTGATCTG

A I Z Z G N E D I A E V L Q K A A

GCAATTTMYTMYGGCAATGAAGATATCGCAGAAGTTCTGCAAAAAGCAGCG

Figure 2. Binary DARPin library sequence. Contains TMY codons (Z) at selected positions leading to random binary amino acid diversity (Serine/Tyrosine). The first residue position of N-cap, D1, D2, D3 and C-cap are bolded and underlined and appear in this order.

Analysis of PelB variant display levels shows that all tested PelB variants exhibited statistically significant improvement in display of the binary library compared to the parental PelB signal sequence with the weakest variant DN10 still being twice as effective at displaying the binary library (Figure 3.). The best PelB signal sequence variant, DN5, exceeded the average display levels of DsbA almost 2-fold and those of PelB parental almost 7-fold. One A8V substitution was found in DN5 sequence. Based on the results from this research and Kulmalas, n-region codon usage might be a major factor in dictating the display efficiency with DN5 better than any other variant, but there being a large gap between it and DN10. PelB variant DH4 was still significant improvement over parental PelB, but barely better than DsbA. The possible additive effect of N- and H-region mutations was analysed by fusing DN5 and DN10 with DH4 into two double mutants (DN5–DH4, DN10–DH4), but no additional benefits were gained. In fact, combining the N- and H-region showed decreased display levels of the binary library compared to DN5 and DH4 alone. DN5–DH4 was still a significant improvement compared to DsbA signal sequence.

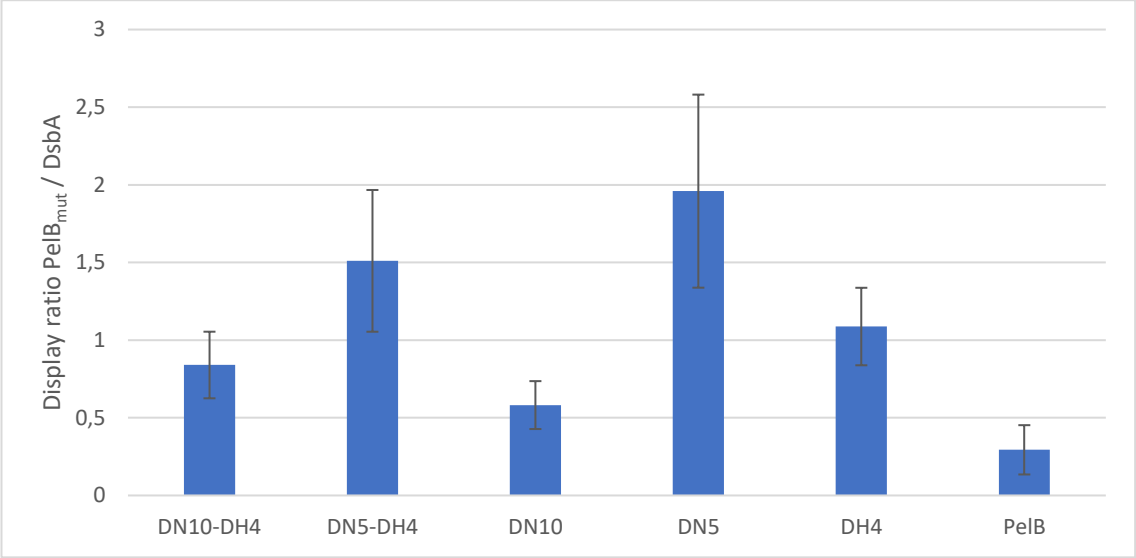

Figure 3. Phage immunoassay of phage PelB signal sequence mutants displaying DARPins with combinatorial Tyr/Ser mutagenesis at paratope. Phage signal is normalized to the DARPin library sample displayed with DsbA. Error bars represent standard deviation of five independent phage productions.

The production of phages and preparation of phage stocks have the greatest potential to bias the results, but I think this is unlikely. The phage production, immunoassays and analysis were done in five replicates to lessen the impact of natural variance. DN5 overperformed DsbA each time and was the top variant in each replicate. There was still variance between replicates as the effectiveness of DN5 could double as seen in the graph above. All the phage titers were normalized using the same DsbA-BiLiBST phage stock as standard, so all phage titers have the same amount of error. After each anti-DARPin immunoassay, a total phage immunoassay was done using the phage dilutions from the anti-DARPin immunoassay. Coefficient of variation (CV%) for phage samples in a plate ranged from 11% to 19%. The cause behind improved display level is not clear, but some studies show that increasing hydrophobicity might change the transportation system from post-translational to co-translational leading to improved phage display (Huber *et al.* 2005). Other signal sequence features could also make the signal sequence SRP dependent. Changes in the basic residues or possible presence of helix-breaking residues in the hydrophobic core might cause a switch to co-translational translocation. Tests conducted by Kulmala show that the alanine to valine substitution could affect the display efficiency positively (Kulmala 2020).

The best identified signal sequence, DN5, was chosen for constructing the main DARPin library using binary DARPin library as basis. Before this could happen, the serine/tyrosine binary library is preselected for open reading frames, because nonfunctional junk genes are a natural consequence of any combinatorial gene synthesis process and frameshifts would decrease the library quality. The library ordered from Eurofins had roughly 50% non-functional clones and is selected for clones with intact frame. Before selection, the serine/tyrosine binary DARPin library was enlarged to over 30 million transformants. The first pEB32D9-BiLiB enlargement yielded 26 million transformants with 1.5% vector background and the second 32 million transformants with 0.9% vector background, giving 58 million transformants in total.

The preselection was performed by panning the larger BiLiBST two rounds against anti-DARPin Fabs. Only clones with intact frames should bind to anti-DARPin Fabs as frameshifts stop the formation of p3 phage surface protein. Two different pannings were conducted side by side, one with just TBT-0.05 as panning buffer and another with added 5% fat free milk to decrease nonspecific binding. After two rounds of panning, the library preselection strategy was confirmed by sequencing 24 clones from both panning reactions. By comparing the sequences together, it can be seen, if the panning had preferred either serine or tyrosine residues or any of the codons coding these two amino acids (Figure 4.) and how much of the frameshifts were removed from the pool. The panning without just TBT-0.05 yielded better results as the binary DARPin library was effectively purified from frameshifts decreasing the frequency of frameshifts decreased from 50 % to 13 %. There was a more even spread amongst the four possible codons and only one non-Ser/Tyr mutation at codon of interest. With added 5% fat free milk, the codon spread was more uneven and there were significantly more non-Syr/Ter mutations at codons of interest. Both pannings preferred tyrosine residues at N-CAP, D1 and D2 and serine residues at D3 and C-CAP.

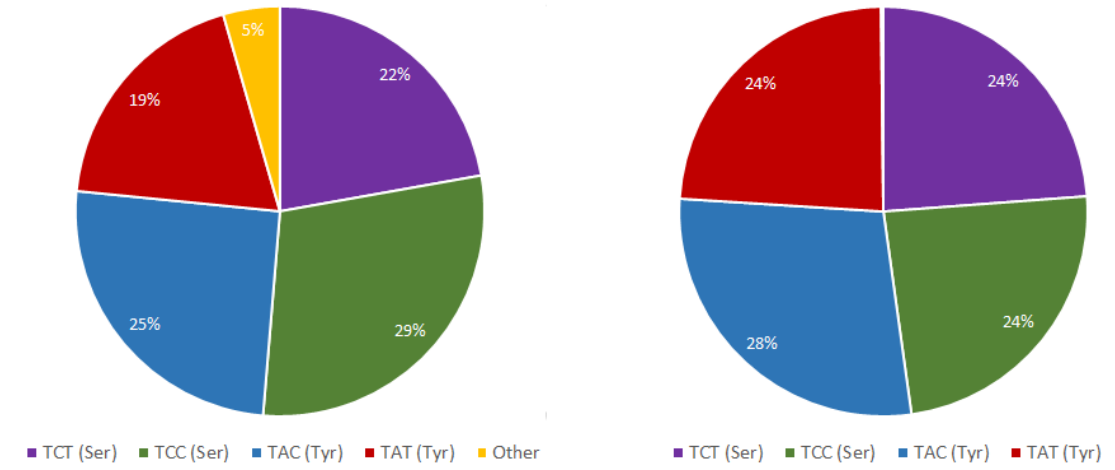

Figure 4. Total codon diversity at codons of interest in binary DARPin library. Leftside has the panning reaction with added 5% fat free milk and rightside has the panning reaction with just TBT-0.05. "Other" marks non-Ser/Tyr mutations at codons of interest.

The phage DARPin library is based on pEB32x vector with modified signal sequence. This post-translational signal sequence was selected in the beginning of research after analyzing display levels of multiple different signal sequences. The main DARPin library was constructed using preselected and enlarged binary DARPin library as template. Construction was conducted by assembly PCR using designed forward and reverse oligos with all the codons of interest replaced in the forward oligo by one of the 18 amino acids, avoiding the residues cysteine (to eliminate disulfide formation) and proline. The end library is comprised of D2 and D3 ankyrin repeat motifs each containing six completely diversified positions flanked by N-cap, D1 and C-cap repeat motifs containing two, six and five Ser/Tyr -diversified positions, respectively. Variable positions are located amongst surface-exposed residues with potential target interactions, while leaving structurally important framework positions fixed. A two-stranded fragment is formed with D2, D3 and C-cap and another fragment with N-cap and D1. This yields a DARPin gene pool with superdiversified central region and minor diversity at N- and C-terminal ends. Assembly PCR was performed with these two fragments followed by SfiI digestion and ligation to phage vector. The main DARPin library (Figure 5.) was constructed in two separate library transformations, libraries DAB B and C, and combined to create ~ 4.9×10^9 cfu library.

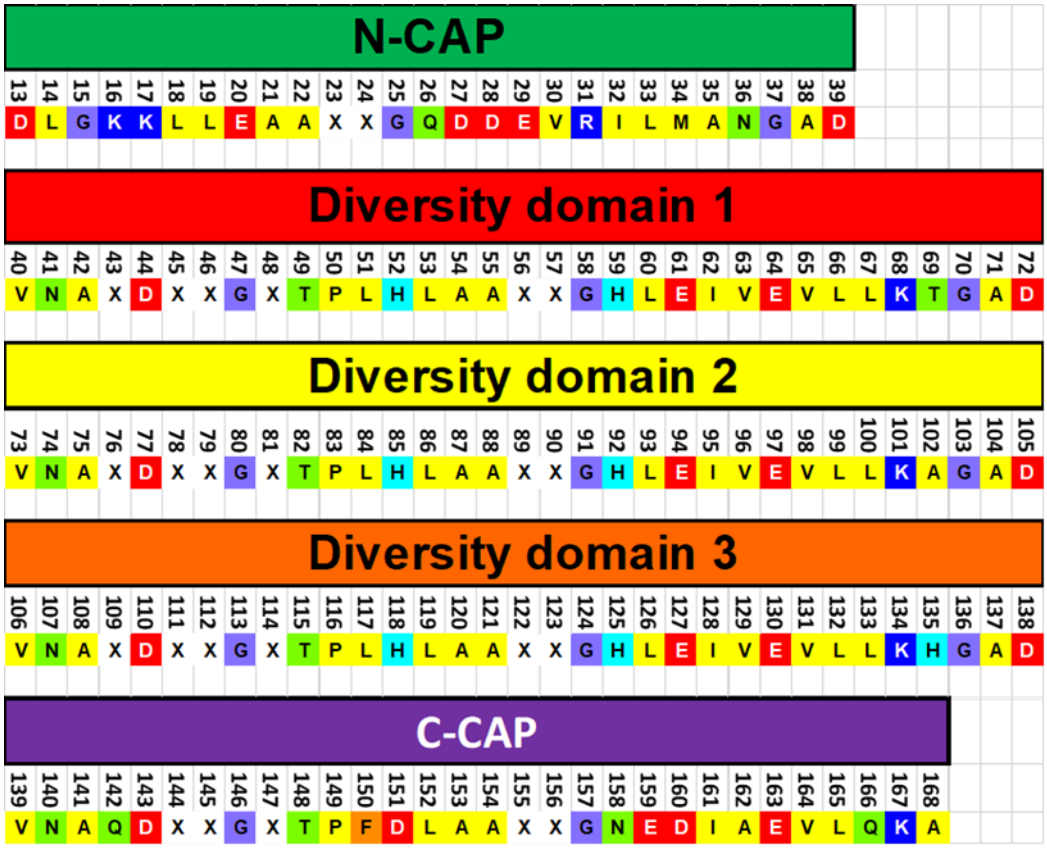

Figure 5. The main DARPin library. Library was diversified by randomization at DARPin diversity domains 2 and 3 by replacing the codon of interest with one of 18 amino acids (cysteine and proline excluded), codons of interest in other domains were left at binary Ser/Tyr diversity. These codons of interest are marked with X.

To validate the main DARPin library, it was selected against biotinylated GST-protein N. The first round of selection was performed on biotinylated antigen bound to streptavidin beads, followed by two further selection rounds in solution with third round having to different antigen concentrations. Before the first and the second selection, the phages were incubated with just GST to direct the protein binder enrichment towards recognizing the nucleoprotein portion. From each selection round alongside non-selected phages the enrichment of binders was analyzed by phage immunoreactivity assay, in which, measured how phages from different selection rounds bound to biotinylated GST-protein N immobilized on streptavidin well surface. The ratio of target binding signal to background was used to measure enrichment. Statistical analysis of immunoreactivity assay results showed increase of binders up to the second round of panning, after which,

it plummeted. The second round had the highest signal to background ratio of 11-fold (Figure 6.).

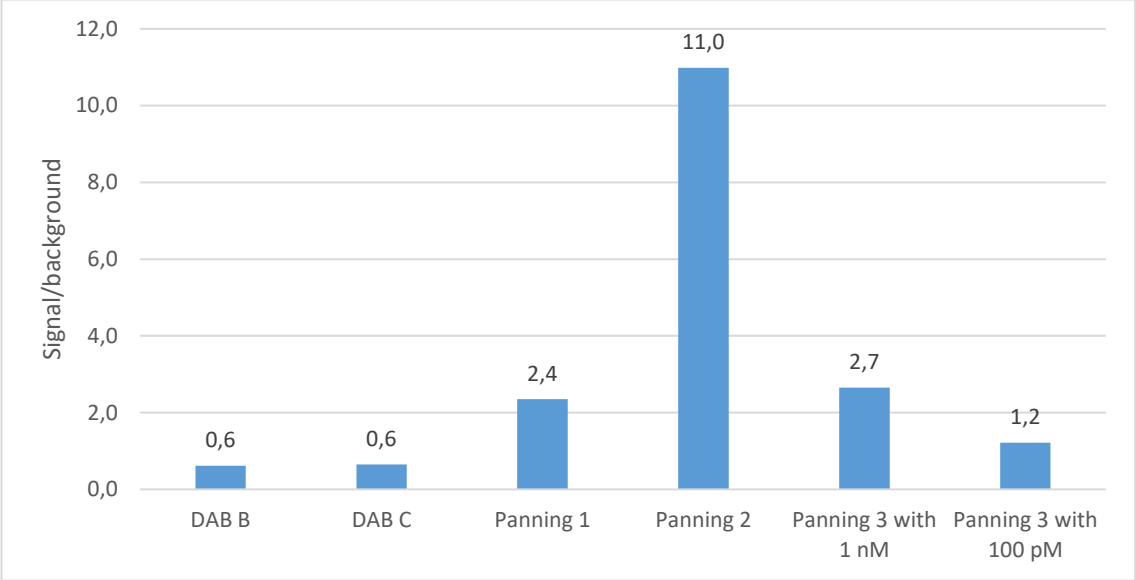

Figure 6. Phage immunoreactivity assay against biotinylated GST-N protein. Depicted is the signal to background ratio of three subsequent selection rounds alongside non-selected phages from DAB B and DAB C libraries.

Even though the signal gets lower on subsequent rounds, there is still clear indication of binding happening at low nanomolar or picomolar antigen concentrations as seen from signal to background ratios and panning outputs (Table 3). Enrichment of phages was already detected during the selections when comparing colony counts of panning reaction and background reaction platings. The use of streptavidin beads all three panning rounds probably explains the drop in signal to background ratio after the second round. Even with all the negative selections performed, it probably was not enough to stop the enrichment of streptavidin binders on meaningful level. The preselection process did not remove all the non-functional of the binary library used that was used as a template and a randomized library naturally has some non-functional parts. These might encode defective molecules with unstructured and sticky polypeptides that bind to background and lead to background enrichment. Antigen concentrations of 1 nM and 100 pM are also really low for testing, increasing potential background. Phage immunoreactivity assay can be used to confirm enrichment of phage particles, but it is too inaccurate to give a

good picture of binders specificity or affinity. More thorough screening process is needed to find out these parameters, which was omitted due to lack of time. To study the functional diversity of the main DARPin library, more selections should be performed against a broad range of target proteins.

Table 3. Panning outputs

Panning round	Antigen concetration (nM)	Output (cfu)
1	62	$3.81x10^5$
2	6.9	$3.40x10^6$
3	1	$1.37x10^5$
3	0.1	$4.21x10^4$

4. Discussion and conclusion

DsbA, which is co-translational signal recognition particle pathway (SRP), is currently the dominant signal peptide used in phage display of DARPins, but even DsbA has a quite poor for display meaning that better signal sequences are needed for higher display efficiency. Closest competitor would be post-translational Sec pathway, which has the advantage of higher transportation capacity. This pathway has, reportedly, highly inefficient phage display of DARPins, but the competing SRP pathway is more likely to overload. PelB was the post-translational Sec pathway of choice, because recently people at University of Turku laboratory of biotechnology managed to improve the expression and Sec-dependent translocation of Fab fragments by selecting improved variants from PelB signal sequence libraries. There is also a number problem associated with phage display as, in theory, full combinatorial mutagenesis of DARPin results in celestial number of variants (1031), but phage display library can have only billion to ten billion unique variants (109–1010). The number problem is solved by ensuring that the binary library used as a starting point is of the highest possible quality by removing frameshifts from it so only open reading frames are left in the library. Furthermore, the randomization is performed in multiple steps. This makes it possible to scan far larger sequence-space than is possible with a single pot library.

The identification of high efficiency PelB signal sequence DN5 for DARPin-phage display and the accomplished first-stage DARPin library with high quality of DNA was a solid foundation for further molecular diversification and the construction of the main DARPin library. Even though the main DARPin library was only selected against one target protein, it already proves the validity of the library and binders detecting picomolar concentration were already enriched. The quality of a combinatorial library is affected by three major factors: library design, theoretical diversity alongside library quality and the chosen display technology. The modified PelB DN5 allows efficient display of DARPins in levels significantly better than the current standard DsbA using post-translational pathway instead of Sec-dependent. Successful enrichment of binders confirm that the library design and phage display enable selection of binding molecules. Even though the main DARPin library size is farcry from the possible diversity that could be reached, it is still large enough to give high enough functional diversity which to select binders from. Further selection should result already in great affinities, but the binders can be further

maturated by changing the flanking Tyrosine/Serine sites (at N-cap+ D1 and C-cap) to fully randomized libraries.

One of the sub goals of this study was constructing the first DARPin library in Finland. From this library, specific binders with high affinities could then be selected against a broad range of target proteins by using phage display. This study is a part of larger research to develop a platform for producing high affinity protein binders. The successful usage of post-translational pathway proves that the co-translational SRP dependent system, like that offered by DsbA, is not the only option to display fast-folding and stable proteins such as DARPins. These finding make phage display a more desirable selection method for DARPin libraries. It can also serve as a foundation for libraries containing other stable and fast-folding proteins.

References

Amstutz, P., Koch, H., Binz, H. K., Deuber, S. A. & Plückthun, A. (2006) Rapid selection of specific MAP kinase-binders from designed ankyrin repeat protein libraries. *Protein Engineering, Design & Selection : PEDS* **19**: 219–229.

Andrade, M. A., Perez-Iratxeta, C. & Ponting, C. P. (2001) Protein repeats: structures, functions, and evolution. *Journal of Structural Biology* **134**: 117–131.

Batyuk, A., Wu, Y., Honegger, A., Heberling, M. M. & Plückthun, A. (2016) DARPin-based crystallization chaperones exploit molecular geometry as a screening dimension in protein crystallography. *Journal of Molecular Biology* **428**: 1574–1588.

Beste, G., Schmidt, F. S., Stibora, T. & Skerra, A. (1999) Small antibody-like proteins with prescribed ligand specificities derived from the lipocalin fold. *Proceedings of the National Academy of Sciences of the United States of America* **96**: 1898–1903.

Bibi, E. (2011) Early targeting events during membrane protein biogenesis in Escherichia coli. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1808**: 841–850.

Binz, H. K., Amstutz, P., Kohl, A., Stumpp, M. T., Briand, C., Forrer, P., Grütter, M. G. & Plückthun, A. (2004) High-affinity binders selected from designed ankyrin repeat protein libraries. *Nature Biotechnology* **22**: 575–582.

Binz, H. K., Bakker, T. R., Phillips, D. J., Cornelius, A., Zitt, C., Göttler, T., Sigrist, G., Fiedler, U., Ekawardhani, S., Dolado, I., Saliba, J. A., Tresch, G., Proba, K. & Stumpp, M. T. (2017) Design and characterization of MP0250, a tri-specific anti-HGF/anti-VEGF DARPin drug candidate. *MAbs* **9**: 1262–1269.

Binz, H. K., Kohl, A., Plückthun, A. & Grütter, M. G. (2006) Crystal structure of a consensus-designed ankyrin repeat protein: implications for stability. *Proteins* **65**: 280–284.

Binz, H. K., Stumpp, M. T., Forrer, P., Amstutz, P. & Plückthun, A. (2003) Designing repeat proteins: well-expressed, soluble and stable proteins from combinatorial libraries of consensus ankyrin repeat proteins. *Journal of Molecular Biology* **332**: 489–503.

Boersma, Y. L. (2018) Advances in the application of designed ankyrin repeat proteins (DARPins) as research tools and protein therapeutics. *Methods in Molecular Biology* **1798**: 307–327.

Boersma, Y. L., Chao, G., Steiner, D., Wittrup, K. D. & Pluc̈kthun, A. (2011) Bispecific designed ankyrin repeat proteins (DARPins) targeting epidermal growth factor receptor inhibit A431 cell proliferation and receptor recycling. *The Journal of Biological Chemistry* **286**: 41273–41285.

Breeden, L. & Nasmyth, K. (1987) Similarity between cell-cycle genes of budding yeast and fission yeast and the Notch gene of Drosophila. *Nature* **329**: 651–654.

Bukowska, M. A. & Grütter, M. G. (2013) New concepts and aids to facilitate crystallization. *Current Opinion in Structural Biology* **23**: 409–416.

Chandler, P. G. & Buckle, A. M. (2020) Development and differentiation in monobodies based on the fibronectin type 3 domain. *Cells* **9.**

Dreier, B., Honegger, A., Hess, C., Nagy-Davidescu, G., Mittl, P. R. E., Grütter, M. G., Belousova, N., Mikheeva, G., Krasnykh, V. & Plückthun, A. (2013) Development of a generic adenovirus delivery system based on structure-guided design of bispecific trimeric DARPin adapters. *Proceedings of the National Academy of Sciences of the United States of America* **110**.

Dreier, B., Mikheeva, G., Belousova, N., Parizek, P., Boczek, E., Jelesarov, I., Forrer, P., Plückthun, A. & Krasnykh, V. (2011a) Her2-specific multivalent adapters confer designed tropism to adenovirus for gene targeting. *Journal of Molecular Biology* **405**: 410–426.

Dreier, B. & Plückthun, A. (2011b). Ribosome display: a technology for selecting and evolving proteins from large libraries. *Methods in Molecular Biology* **687**: 283–306.

Eggel, A., Baravalle, G., Hobi, G., Kim, B., Buschor, P., Forrer, P., Shin, J. S., Vogel, M., Stadler, B. M., Dahinden, C. A. & Jardetzky, T. S. (2014) Accelerated dissociation of IgE:FcεRI complexes by disruptive inhibitors actively desensitizes allergic effector cells. *The Journal of Allergy and Clinical Immunology* **133**: 1709-1719.

Fiedler, U., Ekawardhani, S., Cornelius, A., Gilboy, P., Bakker, T. R., Dolado, I., Stumpp, M. T. & Dawson, K. M. (2017) MP0250, a VEGF and HGF neutralizing DARPin molecule shows high anti-tumor efficacy in mouse xenograft and patient-derived tumor models. *Oncotarget* **8**: 98371–98383.

Forrer, P., Binz, H. K., Stumpp, M. T. & Plückthun, A. (2004) Consensus design of repeat proteins. *Chembiochem : A European Journal of Chemical Biology* ***5***: 183–189.

Galpern, E. A., Freiberger, M. I. & Ferreiro, D. U. (2020). Large Ankyrin repeat proteins are formed with similar and energetically favorable units. *PloS One* **15**.

Gibson, D., Young, L., Chuang, R., Venter, J., Hutchison, C. & Smith, H. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods* **6**: 343–345.

Gilbreth, R. N. & Koide, S. (2012) Structural insights for engineering binding proteins based on non-antibody scaffolds. *Current Opinion in Structural Biology* **22**: 413–420.

Grove, T. Z., Cortajarena, A. L. & Regan, L. (2008) Ligand binding by repeat proteins: natural and designed. *Current Opinion in Structural Biology* **18**: 507–515.

Hanes, J., Jermutus, L. & Pluckthun, A. (2000) Selecting and evolving functional proteins in vitro by ribosome display. *Methods in Enzymology* **328**: 404–430.

Hanes, J. & Plückthun, A. (1997) In vitro selection and evolution of functional proteins by using ribosome display. *Proceedings of the National Academy of Sciences of the United States of America* **94**: 4937–4942.

Houlihan, G., Gatti-Lafranconi, P., Kaltenbach, M., Lowe, D. & Hollfelder, F. (2014) An experimental framework for improved selection of binding proteins using SNAP display. *Journal of Immunological Methods* **405**: 47–56.

Houlihan, G., Gatti-Lafranconi, P., Lowe, D. & Hollfelder, F. (2015) Directed evolution of anti-HER2 DARPins by SNAP display reveals stability/function trade-offs in the selection process. *Protein Engineering, Design & Selection : PEDS* **28**: 269–279.

Houlihan, G., Lowe, D. & Hollfelder, F. (2013) SNAP display - an in vitro method for the selection of protein binders. *Current Pharmaceutical Design* **19**: 5421–5428.

Huber, D., Boyd, D., Xia, Y., Olma, M. H., Gerstein, M. & Beckwith, J. (2005) Use of thioredoxin as a reporter to identify a subset of Escherichia coli signal sequences that promote signal recognition particle-dependent translocation. *Journal of Bacteriology* **187**: 2983–2991.

Interlandi, G., Wetzel, S. K., Settanni, G., Plückthun, A. & Caflisch, A. (2008) Characterization and further stabilization of designed ankyrin repeat proteins by combining molecular dynamics simulations and experiments. *Journal of Molecular Biology* **375**: 837–854.

Kane, E. I. & Spratt, D. E. (2021) Structural insights into ankyrin repeat-containing proteins and their influence in ubiquitylation. *International Journal of Molecular Sciences* **22**: 1–13.

Kohl, A., Binz, H. K., Forrer, P., Stumpp, M. T., Plückthun, A. & Grütter, M. G. (2003) Designed to be stable: crystal structure of a consensus ankyrin repeat protein. *Proceedings of the National Academy of Sciences of the United States of America* **100**: 1700–1705.

Koide, A., Bailey, C. W., Huang, X. & Koide, S. (1998) The fibronectin type III domain as a scaffold for novel binding proteins. *Journal of Molecular Biology* **284**: 1141–1151.

Kramer, M. A., Wetzel, S. K., Plückthun, A., Mittl, P. R. E. & Grütter, M. G. (2010) Structural determinants for improved stability of designed ankyrin repeat proteins with a redesigned C-capping module. *Journal of Molecular Biology* **404**:381–391.

Kulmala A. (2020) UNVEILING THE SECRETS OF DNA: Improved expression and phage display efficiency of synthetic recombinant binding proteins in E. coli through modulation of codon Usage. PhD thesis. 148 p. University of Turku. Turku.

Kulmala, A., Lappalainen, M., Lamminmäki, U. & Huovinen, T. (2022) Synonymous codons and hydrophobicity optimization of post-translational signal peptide PelB increase phage display efficiency of DARPins. *ACS Synthetic Biology* **11**: 3174–3181.

Kummer, L., Parizek, P., Rube, P., Millgramm, B., Prinz, A., Mittl, P. R. E., Kaufholz, M., Zimmermann, B., Herberg, F. W. & 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 of the United States of America* **109**.

Könning, D. & Kolmar, H. (2018) Beyond antibody engineering: directed evolution of alternative binding scaffolds and enzymes using yeast surface display. *Microbial Cell Factories* **17.**

Leemhuis, H., Stein, V., Griffiths, A. D. & Hollfelder, F. (2005) New genotype-phenotype linkages for directed evolution of functional proteins. *Current Opinion in Structural Biology* **15**: 472–478.

Lehmusvuori, A., Tapio, A. H., Mäki-Teeri, P., Rantakokko-Jalava, K., Wang, Q., Takalo, H. & Soukka, T. (2013) Homogeneous duplex polymerase chain reaction assay using switchable lanthanide fluorescence probes. *Analytical Biochemistry* **436**: 16-21.

Li, J., Mahajan, A. & Tsai, M. D. (2006) Ankyrin repeat: A unique motif mediating protein-protein interactions. *Biochemistry* **45**: 15168–15178.

Ling, M. (2003) Large antibody display libraries for isolation of high-affinity antibodies. *Combinatorial Chemistry & High Throughput Screening* **6**: 421–432.

Luo, R., Liu, H. & Cheng, Z. (2022) Protein scaffolds: antibody alternatives for cancer diagnosis and therapy. *RSC Chemical Biology* **3**: 830–847.

Lux, S. E., John, K. M. & Bennett, V. (1990) Analysis of cDNA for human erythrocyte ankyrin indicates a repeated structure with homology to tissue-differentiation and cell-cycle control proteins. *Nature* **344**: 36–42.

Löfblom, J., Frejd, F. Y. & Ståhl, S. (2011) Non-immunoglobulin based protein scaffolds. *Current Opinion in Biotechnology* **22**: 843–848.

Merz, T., Wetzel, S. K., Firbank, S., Plückthun, A., Grütter, M. G. & Mittl, P. R. E. (2008) Stabilizing ionic interactions in a full-consensus ankyrin repeat protein. *Journal of Molecular Biology* **376**: 232–240.

Nangola, S., Minard, P. & Tayapiwatana, C. (2010) Appraisal of translocation pathways for displaying ankyrin repeat protein on phage particles. *Protein Expression and Purification* **74**: 156–161.

Nilsson, B., Moks, T., Jansson, B., Abrahmsén, L., Elmblad, A., Holmgren, E., Henrichson, C., Jones, T. A. & Uhlén, M. (1987) A synthetic IgG-binding domain based on staphylococcal protein A. *Protein Engineering* **1**: 107–113.

Nord, K., Nilsson, J., Nilsson, B., Uhlén, M. & Nygren, P. Å. (1995) A combinatorial library of an alpha-helical bacterial receptor domain. *Protein Engineering* **8**: 601–608.

Pancer, Z. & Cooper, M. D. (2006) The evolution of adaptive immunity. *Annual Review of Immunology* **24**: 497–518.

Parmeggiani, F. & Huang, P. S. (2017) Designing repeat proteins: a modular approach to protein design. *Current Opinion in Structural Biology* **45**: 116–123.

Perez-Riba, A., Synakewicz, M. & Itzhaki, L. S. (2018) Folding cooperativity and allosteric function in the tandem-repeat protein class. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* **373.**

Plückthun, A. (2015) Designed ankyrin repeat proteins (DARPins): binding proteins for research, diagnostics, and therapy. *Annual Review of Pharmacology and Toxicology* **55**: 489–511.

Reichen, C., Hansen, S. & Plückthun, A. (2014) Modular peptide binding: from a comparison of natural binders to designed armadillo repeat proteins. *Journal of Structural Biology* **185**: 147–162.

Richter, A., Eggenstein, E. & Skerra, A. (2014) Anticalins: exploiting a non-Ig scaffold with hypervariable loops for the engineering of binding proteins. *FEBS Letters* **588**: 213–218.

Rothe, C. & Skerra, A. (2018) Anticalin proteins as therapeutic agents in human diseases. *BioDrugs: Clinical Immunotherapeutics, Biopharmaceuticals and Gene Therapy* **32**: 233–243.

Rothenberger, S., Hurdiss, D. L., Walser, M., Malvezzi, F., Mayor, J., Ryter, S., Moreno, H., Liechti, N., Bosshart, A., Iss, C., Calabro, V., Cornelius, A., Hospodarsch, T., Neculcea, A., Looser, T., Schlegel, A., Fontaine, S., Villemagne, D., Paladino, M., … Trimpert, J. (2022) The trispecific DARPin ensovibep inhibits diverse SARS-CoV-2 variants. *Nature Biotechnology* **40**: 1845-1854.

Schilling, J., Jost, C., Ilie, I. M., Schnabl, J., Buechi, O., Eapen, R. S., Truffer, R., Caflisch, A. & Forrer, P. (2022) Thermostable designed ankyrin repeat proteins (DARPins) as building blocks for innovative drugs. *The Journal of Biological Chemistry* **298**.

Schilling, J., Schöppe, J. & Plückthun, A. (2014) From DARPins to LoopDARPins: novel LoopDARPin design allows the selection of low picomolar binders in a single round of ribosome display. *Journal of Molecular Biology* **426**: 691–721.

Schütz, M., Batyuk, A., Klenk, C., Kummer, L., de Picciotto, S., Gülbakan, B., Wu, Y., Newby, G. A., Zosel, F., Schöppe, J., Sedlák, E., Mittl, P. R. E., Zenobi, R., Wittrup, K. D. & Plückthun, A. (2016) Generation of fluorogen-activating designed ankyrin repeat proteins (FADAs) as versatile sensor tools. *Journal of Molecular Biology* **428**: 1272–1289.

Sedgwick, S. G. & Smerdon, S. J. (1999) The ankyrin repeat: a diversity of interactions on a common structural framework. *Trends in Biochemical Sciences* **24**: 311–316.

Sha, F., Salzman, G., Gupta, A. & Koide, S. (2017) Monobodies and other synthetic binding proteins for expanding protein science. *Protein Science : A Publication of the Protein Society* **26**: 910–924.

Siva, A. C., Kirkland, R. E., Lin, B., Maruyama, T., McWhirter, J., Yantiri-Wernimont, F., Bowdish, K. S. & Xin, H. (2008) Selection of anti-cancer antibodies from combinatorial libraries by whole-cell panning and stringent subtraction with human blood cells. *Journal of Immunological Methods* **330**: 109–119.

Škrlec, K., Štrukelj, B. & Berlec, A. (2015) Non-immunoglobulin scaffolds: a focus on their targets. *Trends in Biotechnology* **33**: 408–418.

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

Souied, E. H., Devin, F., Mauget-Faÿsse, M., Kolár, P., Wolf-Schnurrbusch, U., Framme, C., Gaucher, D., Querques, G., Stumpp, M. T. & Wolf, S. (2014) Treatment of exudative age-related macular degeneration with a designed ankyrin repeat protein that binds vascular endothelial growth factor: A Phase I/II study. *American Journal of Ophthalmology* **158**: 724-732.

Speck, J., Arndt, K. M. & Müller, K. M. (2011) Efficient phage display of intracellularly folded proteins mediated by the TAT pathway. *Protein Engineering, Design & Selection : PEDS* **24**: 473–484.

Speltz, E. B., Nathan, A. & Regan, L. (2015) Design of protein-peptide interaction modules for assembling supramolecular structures in vivo and in vitro. *ACS Chemical Biology* **10**: 2108–2115.

Stefan, N., Martin-Killias, P., Wyss-Stoeckle, S., Honegger, A., Zangemeister-Wittke, U. & Plückthun, A. (2011) DARPins recognizing the tumor-associated antigen EpCAM selected by phage and ribosome display and engineered for multivalency. *Journal of Molecular Biology* **413**: 826–843.

Steiner, D., Forrer, P. & Plückthun, A. (2008) Efficient selection of DARPins with sub-nanomolar affinities using SRP phage display. *Journal of Molecular Biology* **382**: 1211–1227.

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**: 823–831.

Steiner, D., Merz, F. W., Sonderegger, I., Gulotti-Georgieva, M., Villemagne, D., Phillips, D. J., Forrer, P., Stumpp, M. T., Zitt, C. & Binz, H. K. (2017) Half-life extension using serum albumin-binding DARPin domains. *Protein Engineering, Design & Selection : PEDS* **30**: 583–591.

Strubel, A., Münick, P., Chaikuad, A., Dreier, B., Schaefer, J., Gebel, J., Osterburg, C., Tuppi, M., Schäfer, B., Knapp, S., Plückthun, A. & Dötsch, V. (2022) Designed ankyrin repeat proteins as a tool box for analyzing p63. *Cell Death and Differentiation* **29**: 2445-2458.

Stumpp, M. T. & Amstutz, P. (2007) DARPins: a true alternative to antibodies. *Current Opinion in Drug Discovery & Development* **10**: 153-159.

Theurillat, J. P., Dreier, B., Nagy-Davidescu, G., Seifert, B., Behnke, S., Zürrer-Härdi, U., Ingold, F., Plückthun, A. & Moch, H. (2010) Designed ankyrin repeat proteins: A novel tool for testing epidermal growth factor receptor 2 expression in breast cancer. *Modern Pathology* **23**: 1289–1297.

Vaccaro, C., Zhou, J., Ober, R. J. & Ward, E. S. (2005) Engineering the Fc region of immunoglobulin G to modulate in vivo antibody levels. *Nature Biotechnology* **23**: 1283–1288.

Valerio-Lepiniec, M., Urvoas, A., Chevrel, A., Guellouz, A., Ferrandez, Y., Mesneau, A., de La Sierra-Gallay, I. L., Aumont-Nicaise, M., Desmadril, M., van Tilbeurgh, H. &

Minard, P. (2015) The αRep artificial repeat protein scaffold: a new tool for crystallization and live cell applications. *Biochemical Society Transactions* **43**: 819–824.

Virnekäs, B., Ge, L., Plukthun, A., Schneider, K. C., Wellnhofer, G. & Moroney, S. E. (1994) Trinucleotide phosphoramidites: ideal reagents for the synthesis of mixed oligonucleotides for random mutagenesis. *Nucleic Acids Research* **22**: 5600–5607.

Walker, R. G., Willingham, A. T. & Zuker, C. S. (2000) A Drosophila mechanosensory transduction channel. *Science (New York, N.Y.)* **287**: 2229–2234.

Walser, M., Rothenberger, S., Hurdiss, D. L., Schlegel, A., Calabro, V., Fontaine, S., Villemagne, D., Paladino, M., Hospodarsch, T., Neculcea, A., Cornelius, A., Schildknecht, P., Matzner, M., Hänggi, M., Franchini, M., Kaufmann, Y., Schlegel, I., Iss, C., Loser, T., … Amstutz, P. (2020) Highly potent anti-SARS-CoV-2 multi-DARPin therapeutic candidates. *BioRxiv.*

Webster, J. M., Zhang, R., Gambhir, S. S., Cheng, Z. & Syud, F. A. (2009) Engineered two-helix small proteins for molecular recognition. *Chembiochem: A European Journal of Chemical Biology* **10**: 1293–1296.

Wetzel, S. K., Ewald, C., Settanni, G., Jurt, S., Plückthun, A. & Zerbe, O. (2010) Residue-resolved stability of full-consensus ankyrin repeat proteins probed by NMR. *Journal of Molecular Biology* **402**: 241–258.

Wetzel, S. K., Settanni, G., Kenig, M., Binz, H. K. & Plückthun, A. (2008) Folding and unfolding mechanism of highly stable full-consensus ankyrin repeat proteins. *Journal of Molecular Biology* **376**: 241–257

Winter, G., Griffiths, A. D., Hawkins, R. E. & Hoogenboom, H. R. (1994) Making antibodies by phage display technology. *Annual Review of Immunology* **12**: 433–455.

Wu, Y., Batyuk, A., Honegger, A., Brandl, F., Mittl, P. R. E. & Plückthun, A. (2017) Rigidly connected multispecific artificial binders with adjustable geometries. *Scientific Reports* **7**.

Wörn, A. & Plückthun, A. (2001). Stability engineering of antibody single-chain Fv fragments. *Journal of Molecular Biology* **305**: 989–1010.

Zahnd, C., Amstutz, P. & Plückthun, A. (2007a) Ribosome display: selecting and evolving proteins in vitro that specifically bind to a target. *Nature Methods* **4**: 269–279.

Zahnd, C., Pecorari, F., Straumann, N., Wyler, E. & Plückthun, A. (2006). Selection and characterization of Her2 binding-designed ankyrin repeat proteins. *The Journal of Biological Chemistry* **281**: 35167–35175.

Zahnd, C., Sarkar, C. A. & Plückthun, A. (2010) Computational analysis of off-rate selection experiments to optimize affinity maturation by directed evolution. *Protein Engineering, Design and Selection* **23**: 175–184.

Zahnd, C., Wyler, E., Schwenk, J. M., Steiner, D., Lawrence, M. C., McKern, N. M., Pecorari, F., Ward, C. W., Joos, T. O. & Plückthun, A. (2007b). A designed ankyrin repeat protein evolved to picomolar affinity to Her2. *Journal of Molecular Biology* **369**: 1015–1028.

Zhang, B. & Peng, Z. Y. (2000) A minimum folding unit in the ankyrin repeat protein p16(INK4). *Journal of Molecular Biology* **299**: 1121–1132.