The Dissertation Committee for Marsha Jane Lewis Certifies that this is the approved version of the following dissertation:

ENGINEERING A PROTEIN FOR PEPTIDE DETECTION AND ALLOSTERIC ACTIVATION

Committee:

George Georgiou, Co-Supervisor

Brent Iverson, Co-Supervisor

Walter Fast

Jennifer Maynard

Pengyu Ren

ENGINEERING A PROTEIN FOR PEPTIDE DETECTION AND ALLOSTERIC ACTIVATION

by

Marsha Jane Lewis, B.S.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

May, 2010

Dedication

This work is dedicated to my loving, accommodating, and supportive husband, Kirk; my inspiration, Ruby Jo; my mother, Louise, who always believes in me; and my late father, Edward, who sparked my fascination with science.

Acknowledgements

I would like to thank Dr. Brent Iverson and Dr. George Georgiou for this graduate research opportunity, your continual support, and believing in me and my projects when I did not.

I thank all BIGG lab members, past and present, for working alongside of me and sharing their productive conversations, particularly Dr. Navin Varadarajan, Dr. Karl Griswold, Dr. Yorgos Skretas, Dr. Everett Stone, Dr. Jack Borrock, Dr. Mridula Rani, Li Zhou, and Mark Pogson. I also thank my undergraduate students, Anna Venardos and Evan Walsh, and graduate rotation student, Tracy Ooi, for their valuable assistance with my experiments.

Finally, thank you to Dr. Ellen Gottlieb, Barbara Welch, and Katherine Reynolds; they made graduate school a reality for me.

ENGINEERING A PROTEIN FOR PEPTIDE DETECTION AND ALLOSTERIC ACTIVATION

Publication No._____

Marsha Jane Lewis, Ph.D.
The University of Texas at Austin, 2010

Supervisors: Brent Iverson and George Georgiou

Strategies for the engineering of allosteric proteins, which are proteins that bind ligands at a specific site other than the reaction site and affect the reaction activity, are still being perfected. There have been allosteric proteins successfully engineered based on the hypothesis that the two allosterically related sites are distinct, modular domains and use trial and error to construct and test novel protein domain fusions for allostery.

This work uses laboratory evolution to engineer the peptide binding affinity of the protein binding domain of the allosteric *E. coli* protease DegS. The protein binding domain is a PDZ domain (named for Postsynaptic density protein (PSD-95), Discs-large protein (Dlg), and Zonula occludens-1 (ZO-1)) that binds the *C*-terminus of unfolded outer membrane porins. Combinatorial libraries of PDZ domain variants were displayed anchored to the periplasmic membrane of *E. coli*. The cells were permeabilized and

incubated with fluorescent peptide ligands. PDZ domains were screened by flow cytometry for binding to the target peptide ligands. The PDZ domain binding affinity was improved by 20-fold for the peptide ligand that represents the physiological ligand; and the PDZ domain binding affinity was expanded to accommodate a negatively charged residue in a novel peptide ligand. The *E. coli* anchored peripalsmic expression (APEx) methodology in conjunction with flow cytometry had not previously been used to modify the binding affinity of a PDZ domain.

The selected PDZ domain variants were then fused to the wild-type DegS protease domain and analyzed to determine if allosteric activation was made more sensitive to the native ligand or altered to respond to the novel peptide ligand. Interestingly, the DegS fusion protein with the PDZ variant containing the most subtle mutations retained a degree of allostery for the physiological peptide ligand and obtained a degree of allostery for the novel activating peptide ligand. Other selected PDZ variants with additional and expected mutations in the ligand binding site did not respond allosterically to the peptide ligands and the respective DegS fusions were constitutively active, suggesting that the amino acid network linking the allosteric binding event to protease activity is intricately integrated.

List of Tables

Chapter 1

Chapter 2

Chapter 3

List of Figures

Chapter 1

List of Figures continued

Chapter 2

Chapter 3

List of Figures continued

List of Figures continued

Appendix I

CHAPTER 1: PROTEIN ENGINEERING AND ALLOSTERIC PROTEINS

INTRODUCTION

Natural selection evolved a diverse and vast array of proteins over billions of years. Natural selection capitalizes on evolutionary processes that occur over time. Protein variants are then selected if they contribute to the organism's increased fitness in the environment. Cycles of DNA mutation followed by environmental selection pressures is repeated over generations. Eventually, these cycles lead to novel proteins and/or functions dependent on the selection pressure.

Natural selection is clearly an effective model for engineering and understanding protein function in the laboratory. However, the goal of laboratory protein evolution is to direct a protein towards a defined, new function within weeks or months versus billions of years. The success of laboratory evolution depends on: 1) sufficient genetic diversity created by DNA mutagenesis, 2) unambiguous selection pressure that is directly related to the desired function, and 3) an accurate screening strategy that efficiently evaluates the desired function of a large number of variants and is linked to the DNA. Choosing an appropriate parent protein which can be evolved to perform a target function is also important to directed evolution [1].

Rational protein design is an alternative to laboratory evolution for the engineering of polypeptides having desired functions. Rational protein design uses available data to formulate and test hypotheses based on the structural information. The success of rational protein design depends on correctly identifying variables that influence the desired function and correctly scaling their impact. A combined approach

that incorporates rational design and laboratory evolution is yet another option for redesigning or probing a protein.

Numerous proteins engineered by rational protein design and/or laboratory protein evolution are currently being applied in biotechnology, diagnostics, and therapeutics. Examples include redesigned ligand-binding proteins such as antibodies displaying increased binding affinity [2-8] and changes in enzyme catalytic efficiency or stability [9-16]. There are also many examples whereby mutations and protein redesign have been used to probe the protein's mechanism of action or to gain insights into protein evolution [17-21].

Strategies for the engineering of allosteric proteins, which are proteins that bind ligands at a specific site other than the reaction site that affect (by inhibition or activation) the reaction [22], are still being perfected although some recent successful examples have been published [23-29]. There are hypotheses that attempt to explain the mechanism by which binding of a ligand onto one site affects the function at a distal site [30-33]. However detailed molecular explanations of allostery are lacking. Engineering allosteric enzymes for activation with non-physiological ligands is an opportunity to further understand some of the fundamental molecular interactions that underlie the mode of action of this very important class of enzymes.

RATIONAL PROTEIN DESIGN

Rational protein design uses available structural and functional information to formulate and test hypotheses iteratively by experiments or computationally. Computational modeling, *in silico* design, is often used in rational protein design to select

variants with the highest likelihood to affect function. The respective variants are synthesized and the effect of the amino acid substitutions on function is determined.

In silico design

In silico design uses known molecular structure information and simulates the effects of amino acid substitutions using computer algorithms. Simulating substitutions computationally was first pioneered in the mid-1980s with a fixed protein backbone template and a set of 67 side chain rotamers assembled after evaluating 2,200 amino acid rotamers in the Protein Data Bank (PDB) [34] from high resolution structures (less than 2.0 angstroms) [35]. Ponder *et al.*'s algorithm evaluated the potential acceptability of an amino acid rotamer primarily based on packing density, van der Waal forces, and steric limitations. The authors did include all the hydrogen atoms but could not include Lys, Arg or Met amino acids due to the high number of potential rotamers and available computational capacity. Since then, algorithms have become more complex and examine a wide range of potential tertiary structure influences simultaneously: including solvent mediated effects [36], electrostatic interactions [37], backbone flexibility, metal binding contributions [38], and typically include, on average, 153 amino acid rotamers [39, 40].

The general approach to *in silico* protein engineering begins with selecting a well-characterized protein template with a solved structure in the PDB [34]. Next, desired amino acids suspected to be relevant for function are selected for substitution. Since it is extremely difficult to computationally quantify the effect of amino acid substitutions on protein function, predictions usually seek to capture the resulting changes in protein conformation and possibly ligand binding. The computer algorithm scores amino acid rotamers in the user-defined positions based on properties that influence tertiary structure stability such as van der Waal forces, hydrogen bonding, solvation

effects and entropy calculations. The best designed protein candidates can then be expressed as recombinant proteins and analyzed experimentally [41].

Sharabi *et al.* sought to demonstrate the effectiveness of computational protein redesign using the algorithm ORBIT [42], to improve the binding affinity between the green mamba snake toxin, fasciculin, and its physiological target, the synaptic enzyme, acetyl cholinesterase. Sharabi *et al.* chose the fasciculin-acetyl cholinesterase model complex because they had a convenient affinity assay for experimental studies; the crystal structure of the complex was known and binding affinity between the two proteins was high, with a binding dissociation (K_D) constant in the sub-nanomolar range. The authors computationally selected thirteen amino acids in fasciculin within 4 angstroms of the molecule's binding interface for modeling by the ORBIT algorithm, which scores rotamer stability within a tertiary structure. The ORBIT designed fasciculin variant (Fas^{des}) contained five mutations. The Fas^{des} bound to the protein ligand with an equilibrium dissociation constant (K_D) measuring 1.2 nM, which was four times higher (i.e. decreased affinity) than the K_D exhibited by the wild-type fasciculin. However, removal of one of the computer predicted mutations, based on inspection of the structure model, ultimately improved binding by decreasing the K_D four-fold compared to wild-type fasciculin [21].

Hypothesis Driven Design

Hypothesis driven design relies on information on the function and structure of a protein but without assistance of computer simulations. For example, in early studies with superoxide dismutase (SOD) [43], an enzyme which prevents oxidative damage by dismuting superoxide radical, O_2^-, to molecular oxygen and hydrogen peroxide, Graf *et al.* hypothesized that substrate diffusion was the limiting factor in catalysis. Based on

inspection of the crystal structure of the enzyme, specific superoxide dismutase residues were mutated to increase the local positive charge and provide electrostatic guidance for the substrate. In this manner, Graf *et al.* successfully improved the catalytic efficiency of SOD.

In hypothesis driven protein design influential amino acids that are distal to the functional site are difficult to predict. A classic example is the conversion of serine protease, trypsin, to a variant with chymotrypsin activity. Trypsin and chymotrypsin are serine protease with similar sequences and tertiary structures. The tertiary structures reveal similarly shaped substrate pockets yet the two enzymes have very different and distinct substrate specificities. Trypsin prefers a basic amino acid in the position *N*-terminal to the cleavage site (P1) while chymotrypsin prefers a bulky hydrophobic amino acid. These two well characterized proteases were ideal as a model to probe the mechanism of substrate specificity. In 1987, it was hypothesized that changing a small set of residues in trypsin to amino acids corresponding to chymotrypsin amino acids in the substrate binding pocket would render the enzyme more chymotrypsin-like in its substrate selectivity [17]. This hypothesis proved incorrect. In fact it took several years of intensive research and fifteen amino acid substitutions to convert trypsin to chymotrypsin activity [18, 19]. It was noted that not all the amino acid substitutions comprised the substrate binding pocket but were also within two distal loops.

LABORATORY PROTEIN EVOLUTION

Laboratory protein evolution is modeled from natural selection. Laboratory evolution typically parallels natural selection in the following ways: (i) a DNA library is first created and transformed into cells, (ii) the cells are subjected to selective pressure

and (iii) desired variants displaying increased fitness are isolated. The genes encoding the selected proteins are then subjected to additional rounds of mutagenesis and screening. As mentioned, the success of laboratory protein evolution depends on: 1) creation of sufficient genetic diversity by DNA mutagenesis, 2) unambiguous selective pressure that is directly related to the desired function and 3) efficient evaluation of the desired function in a large number of variants that are linked to the DNA.

The first step in directed evolution is choosing a template protein coupled to the target function. Next, a genetically diverse pool (a.k.a. library) of mutants is created based on the parent template. A functional mutation is a statistically rare event thus creating a large, diverse number of mutants is necessary for successful protein evolution [1].

Library Construction Methods

Error Prone PCR (ePCR)

ePCR introduces random mutations into a gene during PCR DNA amplification steps. ePCR uses a modified standard PCR protocol and varies the reaction components, including the polymerases, such that DNA amplification errors result. The rate of mutation can be titered by adding the divalent cation, Mn^{2+}, (in addition to the natural DNA polymerase cofactor, Mg^{2+}) and unbalancing nucleotide concentrations [44]. Also, polymerases without proof reading ability are used such as the *Thermus aquaticus (Taq)* polymerase.

ePCR does have an amplification and nucleotide substitution bias and typically incorporates only one substitution within a given amino acid codon. Mutants that are generated early go through more rounds of amplification and are over-represented

creating an amplification bias. Typically, the amplification bias is reduced by combining multiple, parallel ePCR reactions subjected to 10 amplification cycles versus 25 to 35 amplification cycles in standard PCR protocols. The nucleotide substitution bias in ePCR is well-documented for specific polymerases. The substitution bias may be reduced by combining polymerases such as *Taq* and mutazyme that have a complementary bias [45, 46]. The disadvantage of the occurrence of a single base change per codon is that statistically, only five or six alternate amino acids can be encoded at each position. The diversity of an ePCR library can be estimated using statistical programs such as PEDL that take into account the library size, mutation rate, and template length [46].

Oligionucleotide targeted mutagenesis

Targeted DNA mutagenesis requires knowledge of the protein structure and sequence to choose residues for mutagenesis; thus, rational protein design information may be incorporated into laboratory evolution. To implement this method, complementary DNA oligonucleotides are synthesized with random nucleotides in the targeted nucleotide position(s) and used for priming standard PCR amplification reactions. If an equal mixture of the four nucleotides are used during oligionucleotide synthesis then that position in the sequence is represented by "N", where N = A/G/C/T. A codon that is completely randomized is NNN. NNN represents all 64 codons. However, there are three stop codons amongst the 64 codons. Also, some amino acids are encoded by multiple codons (Arg, Leu and Ser) and are over-represented compared to only one codon representing the rarest amino acids, Met and Trp. To avoid stop codons and potential bias for specific amino acids, mixtures of two or three nucleotides are used during oligionucleotide synthesis. The NNB randomization scheme encodes the lower frequency of stop codons, but NNK and NNS randomization schemes result in lower

amino acid bias (S = G/ C, K = G/T, B=C/G/T) (see Table 1). Interestingly, it has been suggested that NNK codons are best for libraries expressed in *E. coli* [47].

Table 1. Statistics of the Randomized Codon Library. The probabilities are calculated based on the genetic code. The randomized codon is represented by an equi-molar mixture of nucleotides as follows: N = A/G/C/T, S = G/ C, K = G/T, B=C/G/T. The percentage of stop codons per codon is the frequency a stop codon will arise. The ratio of the most common amino acid (e.g. serine) to the rarest amino acid (e.g. tryptophan) reflects potential bias for more common amino acids. Data is extracted from [47].

Randomized Codon	Percentage of stop codons per codon	Ratio of probability of most common to rarest amino acid
NNN	5%	6
NNB	2%	5
NNK/NNS	3%	3

Fragment Recombination techniques

ePCR and oligonucleotide targeted mutagenesis methods generate sequence diversity through nucleotide substitutions; however, nature also uses recombination, gene crossover events, to create hybrid DNA sequences. Since 1994, various fragment recombination techniques have been used to create diversity beginning with DNA shuffling [48]. DNA shuffling uses at least two homologous parent genes, which are spliced by DNaseI. Recombination occurs when the templates' complementary regions anneal and are amplified using PCR, which creates a mixture of novel hybrid gene sequences.

Other recombination techniques have been developed to enhance diversity. Briefly, these techniques include the staggered extension process (StEP) [49], random chimeragenesis on transient templates (RACHITT) [50], and iterative truncation for the creation of hybrid enzymes (ITCHY) [51]. Homologous recombination techniques include StEP and RACHITT, in addition to DNA shuffling and are illustrated in Figure 1. Homologous recombination techniques require at least 60% sequence homology between the parent templates [52]. StEP uses a shortened annealing/extension cycle in the PCR

cycle versus splicing parental templates. RACHITT attempts to reduce template bias by using the least homologous protein as the template. ITCHY is one of several methods introduced to eliminate the need for sequence homology between the parent templates. In ITCHY, each gene is truncated from opposite ends by different, complementary exonucleases in a time-dependent manner followed by creating blunt ends with restriction enzymes. The truncated sequences are then ligated together [51].

There are many variations of these fragment recombination methods, but the end product is a mixture of shuffled genes containing sequences from one or more of the parent templates. Laboratory evolution assumes that some of these novel genes are functional.

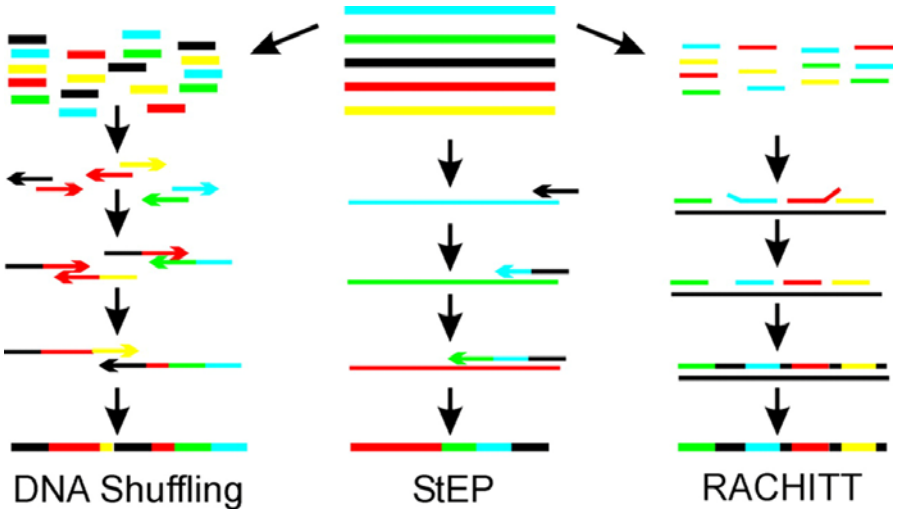


Figure 1. Homology based methods for recombining DNA sequences.
These fragment recombination DNA shuffling methods, DNA Shuffling, StEP, and RACHITT, use at least two homologous parent proteins. The end product is always a pool of hybrid genes containing fragments of the original parental sequence. This figure is reproduced with permission from Oxford University Press [53].

Selection and Screening Strategies

The next step in laboratory protein evolution is to express the library of mutants in a host organism and apply a selective pressure with a selection or screening strategy to isolate desired variants. The selective pressure needs to be specific and can include a counter-selection pressure, if it is also desired to remove or minimize a particular function while also enhancing a different one. A screening strategy involves evaluating each variant individually. A selection strategy will evaluate many variants at one time and permit the positive variants to survive.

All selection and screening strategies have the following attributes: 1) phenotype, the functional trait, and genotype, the DNA sequence, are linked, 2) variants or mutants are subjected to precise selection pressure for the desired trait, and 3) the selected variants or mutants can be amplified or isolated for further evaluation. It is also important that the strategy is able to evaluate a large number of variants or mutants to take advantage of a large, diverse genetic library of mutants, i.e. high throughput.

In general, selection strategies are categorized as cell-based or cell-free, *in vitro*. Cell based assay throughput is limited by the DNA transformation efficiency of the host organisms which can be up to10^{11} for gram negative bacteria [54]. *In vitro* assay throughput is limited by the DNA library construction method and can the order of 10^{15} mutants [55].

Cell based screening strategies will be described further and can be sub-categorized by where the variants are expressed in the cell as “surface display methods” or “display-free methods”. Many cell based strategies use the cell surface as a display platform and enable the expression of the protein variant on an externally accessible cell surface for subsequent screening. Other cell based strategies express the variants in the cytoplasm and are referred to as “display- free”.

Cell based Display-free Methods

Cell based display-free methods use conventional microbiology techniques to engineer or investigate enzyme activity and protein binding domains. These screening methods are simple, relatively inexpensive, and typically do not require specialized equipment or instrumentation for screening the variants.

Genetic selections More often, genetic selections are used for engineering enzymes that have the potential to parallel a physiological enzyme, usually in a metabolic pathway such as amino acid synthesis. The target enzyme is mutagenized and transformed into a cell strain with a corresponding non-functional physiological enzyme. The cells are grown in media without the amino acid, for example, thereby evolving the activity of the target enzyme to complement the deficiency resulting from the non-functional physiological enzyme [10, 56]. This selection method evaluates a large number of cells at once with the surviving cells isolated in the growth media. Surviving cells are expected to contain successful variants; however, this is not always the case as other endogenous metabolic pathways or enzymes can also evolve and compensate since microbial genomes are designed to adapt rapidly to environmental conditions [57].

Yeast Two-hybrid method. The yeast two-hybrid method, developed in 1989, exploits a natural phenomenon in yeast where modular transcriptional activators dimerize to initiate DNA transcription [58]. A binding domain (BD) and an activating domain (AD) comprise the transcriptional activator dimer. The “bait” protein is expressed as a fusion with the BD and the “prey” is expressed as a fusion with the AD. If the “bait” and “prey” interact, the reporter gene is expressed (Figure 2) [58]. Reporter genes are typically enzymes that are required for metabolism to complement deficient media or

catalyze a chromogenic substrate for visual confirmation of positive clones. The assay is sensitive to weak interactions since a single interaction is amplified.

The yeast two-hybrid system has been applied extensively to investigate protein-protein interactions [59-62]. Based on the success of the yeast two-hybrid assay isolating specific protein interactions, the yeast two-hybrid system has been applied to engineer protein binding domains to novel peptide ligands [63, 64]. The yeast two-hybrid strategy used by Schneider *et al*. and Junquiera *et al.* for engineering protein binding domains for a novel peptide ligand included expressing a mutant library of protein binding domains fused to the yeast GAL4-AD (transcription activator GAL4-activating domain) and the novel peptide was fused to the yeast GAL4-BD (GAL4 DNA binding domain); when the two domains dimerized the complex regained its ability to activate transcription from promoters containing Gal4 binding sites, which included the GAL1 promoter fused to the *lacZ* gene which encodes β-galactosidase protein. Positive mutants are detected using x-gal (5-bromo-4-chloro-3-indolyl-b-D-galactopyranoside), a chromogenic substrate of β-galactosidase that is cleaved into galactose and 5-bromo-4-chloro-3-hydroxyindole. 5-bromo-4-chloro-3-hydroxyindole is oxidized into an insoluble visible blue product. Positive clones are then further characterized by more quantitative methods to determine binding affinity. Generally, the drawbacks that limit the utility of the yeast two-hybrid method include the proteins must be expressed in the cytoplasm and transported to the nucleus, thus limiting it to proteins able to fold in a reducing environment, and, most significantly, false positives are common.

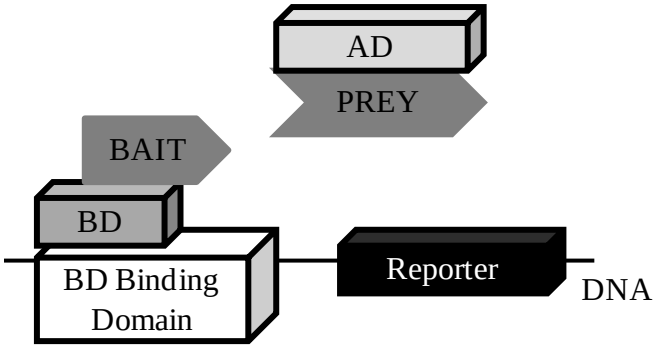


Figure 2. Schematic of the yeast two-hybrid system.
BD is fused to the "bait" protein of interest and binds to DNA binding domain within the promoter region of the reporter gene. AD is fused to the "prey", often a library of potential interacting proteins. When the "bait" and "prey" dimerize, transcription of the report is activated. The reporter is constructed to permit a selection.

The yeast two-hybrid system has been modified for specific cases to reduce false positives [65, 66]; still, the more robust and flexible cell display methods are more often used to engineer protein binding domains.

Cell based Display Methods.

Phage Display. Since the development of phage display in 1985 [67], it has been used to increase stability, expression, and binding affinity of antibodies [68, 69] and non-antibody proteins [70] as well as for enzyme screening [14, 71]. The M13 bacteriophage is most often used for phage display. Briefly, the DNA library is cloned into a phagemid vector, which is a vector with phage and plasmid *ori* sites, and fused to a secretion signal sequence and a portion of one of the five phage coat proteins. The phagemid is transformed and amplified in *E. coli*. *E. coli* is co-infected with helper phage, which then prompts phage particle formation and cell lysis leaving the protein of interest expressed as a phage coat protein fusion on the phage surface. Importantly, this phase contains the DNA encoding the displayed protein

Selecting the desired protein variants displayed on the phage particles is called panning. The phage particles are incubated with an immobilized antigen or peptide. The

phage particles that do not bind are washed away. Bound phage particles are eluted by disrupting the binding without compromising the phage infectivity. These selected phage particles are recycled and amplified in *E. coli* (see Figure 3). Phage display is often applied to redesign proteins for new or improved binding function and also to isolated ligand binding proteins [72-74].

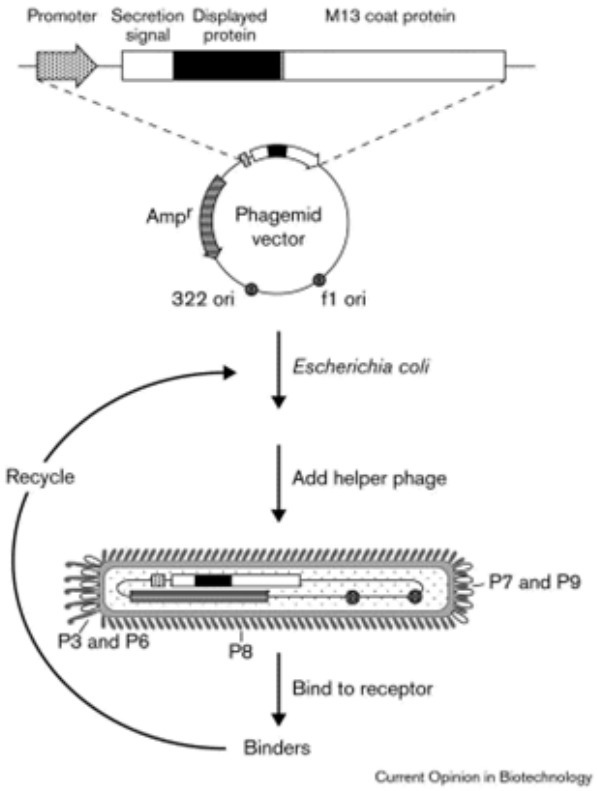


Figure 3. An M13 phagemid vector designed for phage display.
A phagemid vector contains origins of single-stranded (f1 ori) and double-stranded (322 ori) DNA replication and a selective marker, such as the β-lactamase gene (Ampr) which confers resistance to ampicillin. For phage display, the phagemid also contains a cassette consisting of a promoter that drives transcription of an open reading frame encoding a secretion signal and the displayed protein fused to an M13 coat protein. The vector replicates in E. coli as a double-stranded plasmid, but co-infection with a helper phage results in the production of single-stranded DNA that is packaged into phage particles. The phage coat contains five different proteins, and polypeptides can be displayed as either amino-terminal fusions (with P3, P8, P7, or P9) or c-terminal fusions (with P6, P8, or P3). The phage particles can be used in binding selections, and binding clones can be amplified by recycling through an E. coli host. This figure is reproduced with permission from Elsevier [75].

Microbial cell display. The two most common microbial cell display methods use gram negative bacteria *E. coli* and yeast *S. cerevisiae* as the host organisms due to their high transformation efficiencies and fast growth rate. Gram positive bacteria have lower transformation efficiencies, often excrete extracellular proteases, and are not as well characterized genetically as gram negative bacteria [76]. Mammalian or insect cells are used as expression hosts when specific post-translational modifications are required [77].

The first general step of displaying a protein variant on a microbial cell surface is to clone the DNA mutant library into a vector as a fusion with a localization signal sequence and a secondary surface anchoring protein, if the protein of interest is not normally anchored on the cell surface. The vectors are transformed into the cells. The transformation efficiency is the limiting parameter of library size. *E. coli* transformation efficiency is 10^{11} and for *S. cerevisiae* it is typically 10^{7} [54]. The drawback of using the panning process used in phage display is the impact of avidity interactions, which result when the protein is highly expressed on the surface or sufficiently large that it permits the ligand to crosslink multiple surface displayed proteins [54]. The avidity interactions cause the binding affinity to appear improved by decreasing the apparent dissociation rate between the ligand and the protein displayed on the cell surface. The avidity effect with panning is circumvented by screening with flow cytometry using fluorescent ligands or substrates for cell surface display and setting sorting parameters to ignore aggregates of cells.

Flow cytometry screens up to 10^{4} cells per second without loss of cell viability. In flow cytometry, cells are injected into a stream of sheath fluid and forced into single-file order by hydrodynamic focusing. The cells emerge from a nozzle in a droplet. Single cells are analyzed by a detector, which measures the scattering and fluorescence emission after they pass a focused laser beam. If the cell meets the set criteria, the droplet

is electrically charged and deflected into a specific collection container by a strong electric field. Up to five criteria can be screened simultaneously. This is called fluorescent-activated cell sorting (FACS) and depicted schematically in Figure 4 [78].

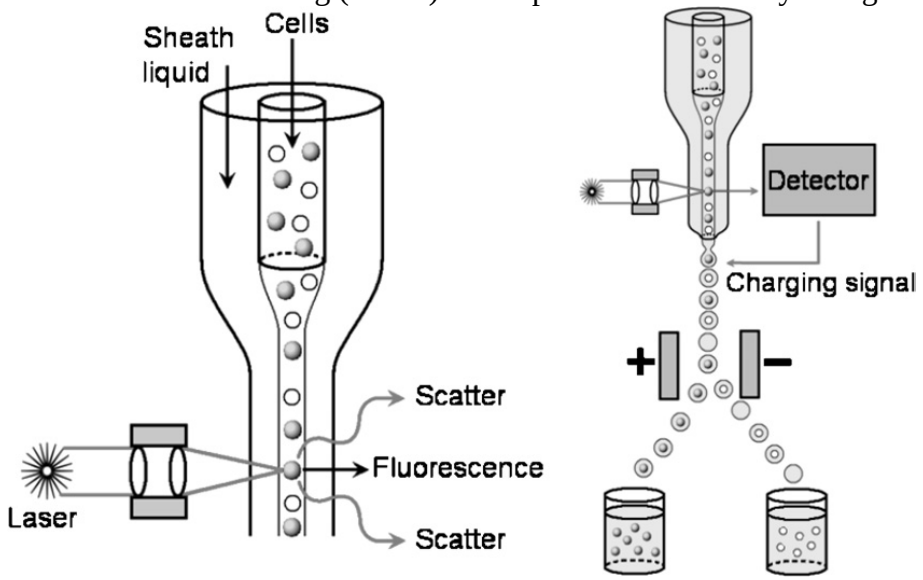


Figure 4. Schematic of flow cytometry and fluorescence-activated cell sorting (FACS) On the left, cells are injected into the sheath liquid and forced to a narrow stream. As fluorescent labeled cells flow past a focused laser beam, they generate light scattering and fluorescence emission measured by optics and electronics. On the right, following optical detection of cells, the mechanical vibration of a nozzle tip generates droplets encapsulating cells. The droplets are selectively charged depending on the presence of labeled cells and deflected by an electric field to different collectors. Figure used with permission from IOPP [78].

E. coli cell display techniques have used the outer and inner cell membrane and the periplasmic space to display the target protein for engineering. Initial efforts displayed the target protein as a fusion with the extracellular loop of outer membrane proteins such as PhoE [79] and LamB [80]. However, the target protein size and folding mechanism limited the application of these methods. An improved outer membrane display method was later designed by fusing the target protein with a localization signal sequence (the first nine amino acids of Lpp) and five transmembrane loops of the integral outer membrane protein OmpA (see Figure 5.A). This fusion was used to successfully

display an enzyme, β-lactamase, and a digoxin single chain antibody onto *E. coli* [81, 82]. The method was proved to be a potential tool for laboratory evolution by successfully recovering active digoxin binding antibodies from a 100,000:1 mixture of non-binding antibody to digoxin binding antibody after incubation with fluorescently labeled digoxin hapten and sorting by FACS [81]. Later, this method was applied to successfully engineer single chain digoxin antibodies for higher digoxin hapten affinity three-fold over the wild-type single chain digoxin single chain antibody fragment [83, 84] A drawback of *E. coli* outer membrane display is the outer membrane negative charge can potentially interfere with protein interactions.

In PECS (periplasmic expression with cytometric screening), the desired protein is secreted into the periplasmic space. To allow access of the desired protein to a ligand, the outer membrane needs to be selectively permeabilized. This can be done by incubating the cells in hyperosmotic buffers. The outer membrane has to be permeabilized in such as way as to balance cell viability while maximizing diffusion of fluorescent ligand to the periplasm without permitting the protein to diffuse away from the cell (see Figure 5B) [85]. PECS was used to improve the equilibrium binding dissociation constant (K_D) of a digoxin single chain antibody with a fluorescent digoxin hapten from 0.9 mM to 150 pM after two rounds of FACS with a library of 10^6 variants. The disadvantage of PECS is the size limit of the fluorescent interacting ligand to 10 kDa.

The next development in *E.coli* display methods was anchored periplasmic expression (APEx) with cytometric screening [86]. The protein targeted for selection is anchored to the periplasmic side of the inner membrane (Figure 5C) as an *N*-terminal fusion to a transmembrane protein such as the six-residue sequence of lipoprotein NlpA, or as a *C*-terminal fusion to the phage gene III minor coat protein of M13, which also

provides the advantage of permitting phagemid libraries to be used in APEx without further DNA manipulation. The size limitation of the fluorescent interacting protein is eliminated in the APEx method by converting the cells to spheroplasts, which are spherically shaped cells due to membrane tension after removing the more rigid cell wall with chemical and enzymatic degradation. The fluorescent ligand is incubated with the spheroplasts followed by screening with FACS. The cells are no longer viable after degrading the outer membrane and the selected spheroplasts are recovered by PCR amplification. During PCR amplification, additional mutagenesis can be done before transforming mutants back into *E. coli* for subsequent FACS screening.

At this time, APEx has been used to increase the binding affinity of multiple single chain and full length antibodies [5, 86-88]. Recently, APEx was applied to screen for antibodies that could bind in a reducing environment. Antibodies rely on di-sulfide bonds for stability. In a reducing environment, antibodies cannot make di-sulfide bonds and thus can neither fold correctly nor bind to antigen. Seo *et al.* redesigned a single chain antibody using an ePCR mutant library with 10^7 mutants. The cells were screened by FACS using a fluorescent antibody to label the bound antigen that was simultaneously expressed with the mutant protein. They isolated two variants that could bind antigen in a reducing environment [89].

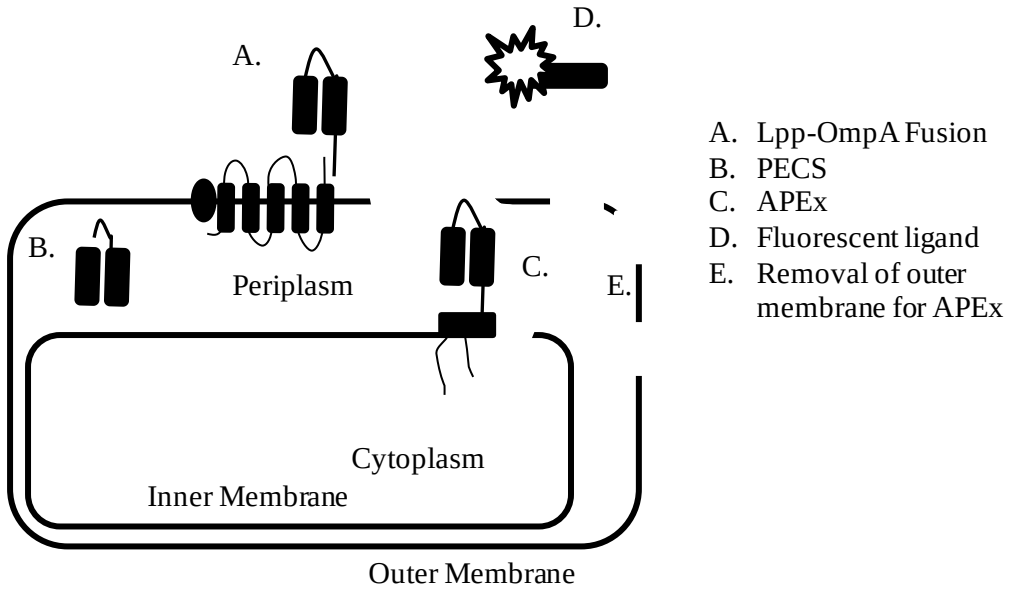


Figure 5. Summary of *E. coli* bacterial cell surface display strategies. A.) An outer-membrane display method developed by this lab using an Lpp-OmpA fusion. B.) Periplasmic expression with cell sorting (PECS), employs the "display" of protein in the periplasmic space. Selective permealization of the OM is employed to allow fluorescent liagdns to equilibrate into the periplasmic space and bind to the target protein. C.) The anchored periplasmic expression system (APEx) with cell sorting, anchored the target protein with different fusions to the periplasmic membrane. In each method FACS was used to screen for variants that bound D.) fluorescent ligand. The anchored protein in APEx was exposed for ligand binding after E.) the outer membrane was degraded.

S. cerevisiae cell surface display was developed after *E. coli* cell surface display in 1997 [90]. In this system proteins are displayed on the yeast cell surface as a fusions to the α-agglutin2 yeast adhesion receptor [91]. The protein can also be fused to an additional protein sequence that can be detected by commercially available antibodies labeled with a fluorophore. As a result, a cell protein can be screened for expression and binding affinity simultaneously (Figure 6). Boder *et al*. constructed libraries of 4-4-20 single chain antibody mutants with an adaptation of DNA shuffling and random mutagenesis methods. The libraries were screened and sorted using FACS. This strategy increased the antibody binding affinity to fluorescein from 0.31 nM to 48 fM, the highest engineered binding affinity reported for a protein [3].

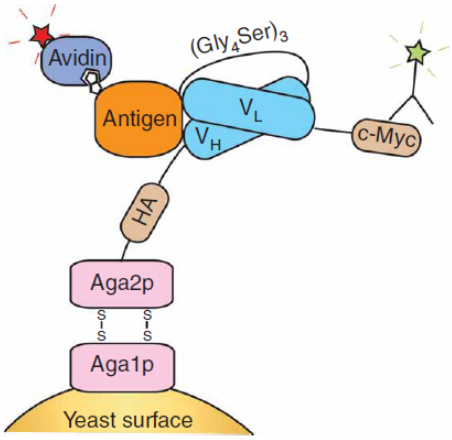


Figure 6. Yeast surface display. The single chain antibody (cyan) is displayed as an Aga2 (pink) fusion protein on the surface of yeast. Expression can be detected by using fluorescent antibodies binding to the epitope tags (beige), and binding of the scFv to a biotinylated antigen (orange) can be detected using fluorescent avidin (violet). HA, hemagglutinin; VL, variable light chain; VH, variable heavy chain; (Gly4Ser)3, flexible peptide linker. This figure and caption are used with permission from *Nature Protocols* [92]

PROTEIN BINDING DOMAINS

At this time, antibody binding domains have been the most commercially successful protein binding domain used to engineer proteins with novel or improved binding specificity and selectivity for biotechnology and therapeutic applications. There are approximately 10,000 antibodies commercially available for biotechnological applications such as protein purification and detection [6]. In 2009, there were twenty-two therapeutic monoclonal antibodies on the U.S. market with eight of them reaching $1 billion in sales [93]. There were also several hundred drug candidates in clinical development in 2005 [4]. Nevertheless, antibodies also have limitations. They are expensive to manufacture and not necessarily stable in reducing environments, as they rely on disulfide bonds for proper folding and stability [94].

Alternative protein binding domains exist in nature as part of signaling complexes and cellular infrastructure. Over the last two decades, many of these protein binding domains have been the target of protein engineering to investigate their function and commercial potential. An excellent comprehensive review describes 43 non-antibody protein binding domains that have been engineered for binding affinity [94]. Another review indicates there are over 50 protein binding domains that have been subjects of protein engineering [95]. In this investigation, a PDZ protein binding domain template was used for engineering binding affinity toward a peptide ligand.

PDZ PROTEIN DOMAINS

Background

The PDZ domain is named for the three proteins where it was initially recognized: postsynaptic density protein (PSD-95), discs-large protein (Dlg), and zonula occludens-1 (ZO-1) [96-99]. Up until 1996, it was thought that PDZ domains were only represented in vertebrates and invertebrates. However, it was since discovered that PDZ domains exist in plants, bacteria and yeast [100]. The ubiquitous PDZ domain is comprised of 80 to 100 amino acids and typically consists of six β strands folded into a β sandwich and two α helixes. The accepted general PDZ numbering scheme is based on the conserved secondary structure elements. For example, αB:3 refers to the third residue in the second conserved α helix. The ligand binding groove is between the groove of the βB strand (the second conserved β strand) and αB helix with a conserved hydrophobic sequence formed by the loop between βA and βB strands that binds the carboxy terminus of the peptide ligand [101, 102]. The peptide ligand *C*-terminal residue is referred to as position (0), followed by position (-1) and so on. The nomenclature scheme is illustrated in Figure 7.

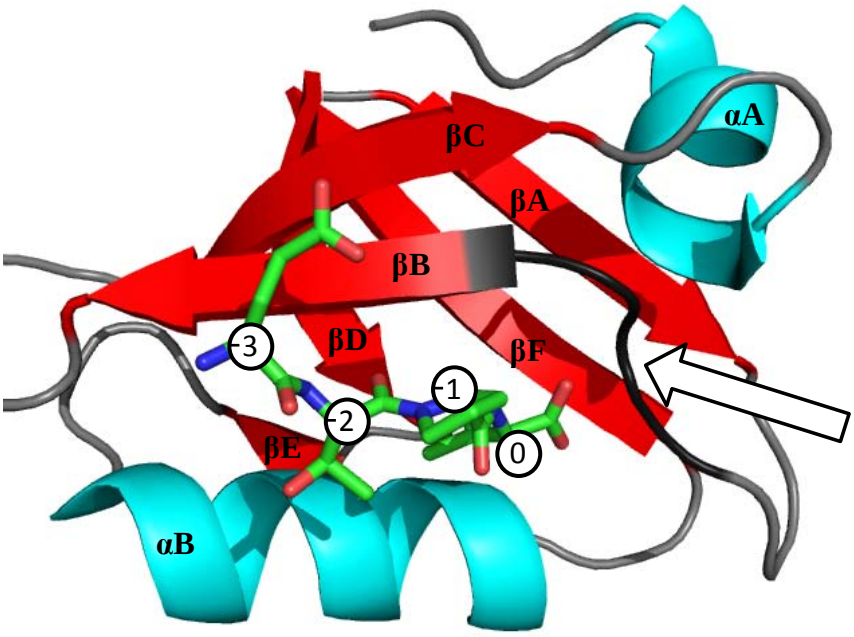


Figure 7. Structure of PSD-95/3 bound to ligand peptide illustrating PDZ domain nomenclature used in this dissertation (PDB:1TP3 [103]). The α-helix structures are in blue and labeled alphabetically from N- to C-terminus. The β-strand structures are in red and also labeled alphabetically from N- to C-terminus. The peptide ligand is labeled with "0" indicating the C-terminal residue. The conserved hydrophobic ligand binding pocket, represented by GLGF in this protein, is highlighted in black in the loop between βA and βB and indicated with a block arrow. This diagram was generated using PyMol [104].

PDZ domains have diverse specificities. PDZ domains predominantly bind the *C*-terminus of proteins, recognizing at least the terminal three residues, but they can also bind internal motifs, lipids, and other PDZ domains [105-108]. PDZ domains often serve as mediators in molecular complexes, for example in clustering functionally relevant molecules in neuronal and immunological synapses, tight junctions, mediating adhesion, and ion transport [109-118]. PDZ domains play important roles in disease pathology including viral exploitation of cells. For example the high risk human papilloma virus expresses an early stage oncogene, E6, with a PDZ binding motif. The general role of the E6 protein is to mediate cell proliferation by disrupting apoptotic cell signaling. Specifically, high-risk E6 proteins bind with the PDZ domains of hDlg (human homologue of Dlg, discus large protein) and hScrib (human homologue of *Drosophila*

melonagastar Scrib, scribble protein) that are hypothetically involved with regulation of cell growth and adhesion. [119-128].

Probing PDZ ligand specificty

There have been many attempts to organize PDZ domains into subgroups or classes. Organizing PDZ domains into related groups can facilitate rational engineering and improve the understanding of the PDZ ligand binding mechanism. It is clear the PDZ family members are related due to their homologous secondary structure elements and a conserved four residue hydrophobic sequence; however, assembling them into subgroups is difficult. Initially, PDZ domains were grouped by their preferred ligand binding sequence. Ligand sequence nomenclature is represented by the Seefeld Convention 2001 nomenclature [129], specifically with X denoting any L-amino acid, Φ representing hydrophobic amino acid residues, Ψ aliphatic amino acid residues.

In early studies, Sonyang *et al.* passed a soluble mixture of 8mer peptides through a column of beads with an immobilized PDZ domain. The peptides retained were then sequenced to determine a consensus ligand sequence. After analyzing nine PDZ domains, they determined two classes of PDZ domains existed based on the three terminal residues of the bound peptides: Class I, which bound S/T-X-Φ ligand sequence; and Class II, which bound Φ-X-Φ ligand sequence [102].

Numerous PDZ domains were discovered and characterized between 1997 and 2001 and many of them did not bind ligands that permitted them to be classified as Class I or II [118, 130]. A Class III group was added to include the group PDZ domains binding ligands with the D/E-X-φ consensus sequence. Bezprozvanny attempted to resolve the classification system by hypothesizing two residues in the PDZ domain that comprise the ligand binding pocket, the last residue of βB and first residue of αB,

determine ligand specificity. He proposed 25 PDZ sub-groups based on potential amino acid combinations in these two positions and the PDZ domains' ligand binding specificity as determined by yeast two-hybrid screening. Despite the extensive group list, there were still some PDZ domains in more than one group [131]. Although, additional classes were named and still referred to today: Class III (D/E-X-φ), a "novel" Class III (-E/D-X-W-C/S) and Class IV (-X-Ψ-D/E) [132, 133]; this overall approach did not catch on but instead prompted further attempts to classify PDZ domains.

As more crystal structures of PDZ domains bound to the ligand were elucidated it was suggested that PDZ domains could be classified by structure based features related to peptide recognition as opposed to only the amino acid sequence of the peptide ligand [134]. At least six different classification categories could be defined based on available structures and sequences. Again, some individual PDZ domains were classified in more than one category and this idea did not become popular either. However, this work resulted in an excellent compilation of different binding strategies, and demonstrated the binding flexibility within the PDZ family (Figure 8).

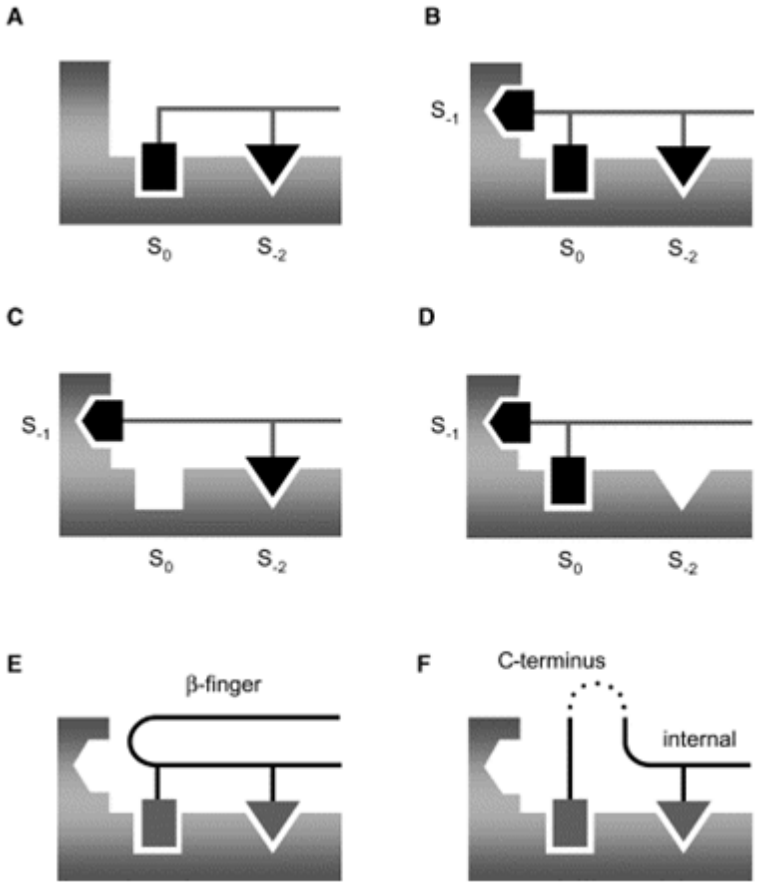


Figure 8. Schematics of PDZ interactions.

A) Classic *C*-terminal sequence recognition B) Three *C*-terminal residues are involved in binding C) Syndecan PDZ2 peptide binding complex D) Syntenin PDZ2-interleukin 5 receptor binding complex E) Syntrophin PDZ binding complex F) Interaction with an internal residue as seen with syntenin PDZ2-PDZ2 complex. This figure and caption is reproduced with permission from Elsevier [134].

In 2008, Chen *et al.* published a statistical model mapping the importance of each PDZ position relative to ligand residue [135]. The model was based on microarray data they had accumulated from analyzing 157 mouse PDZ domains. 16 residues were found on the PDZ domain (1-16) that could be linked to the last four residues of a ligand (Figure 9). Their methodology concurred with existing data noting the first position in αB was related to position -2 on the peptide ligand. However, the broader conclusion

was that a single residue on the PDZ domain recognized more than one ligand residue and multiple residues on the PDZ domain were required to recognize each ligand residue. Thus, there is much overlap amongst the PDZ residues binding to peptide ligand that creates a complex binding mechanism and not a simple "lock and key" or even "induced fit" mechanism.

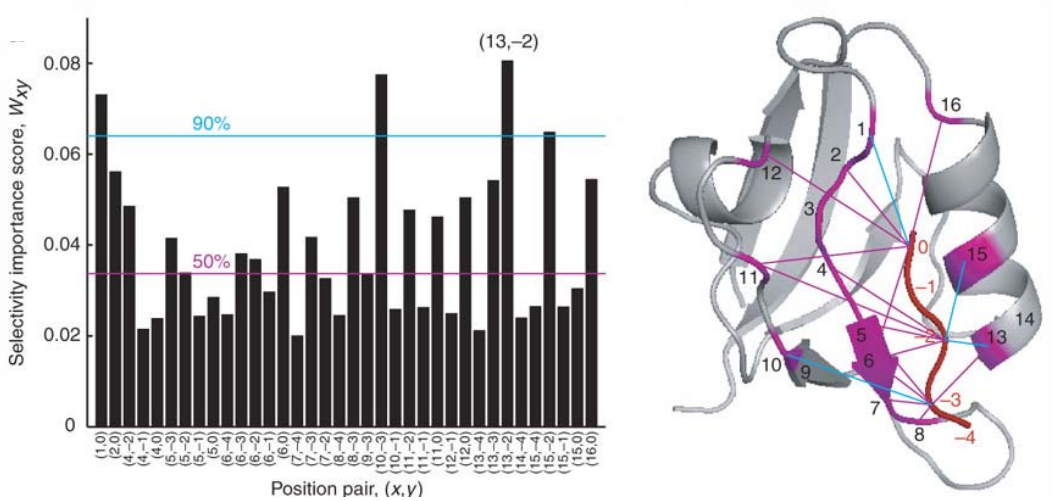


Figure 9. Position pairs that predict peptide-binding selectivity between the PDZ and ligand.
(A) On the x axis: position x = the residue notated on the PDZ domain (B) on the right and position y = the ligand residue. Position pairs (3,0) and (3,-1) were excluded due to high conservation at position 3 (normally a Gly). The magenta line represents the median score of the pairs and the cyan lines represent the 90th percentile. This figure is reproduced with permission from *Nature Biotechnology* [135]

The most recent and comprehensive study linking the PDZ domain's primary sequence to binding specificity was reported by Tonikian *et al.*, who approached the subject with a brute force approach involving the cloning and expression of 88% of the PDZ domains in *C. elegans* and 39% of the PDZ domains from *H. sapiens*. Of the 168 clones, they purified 87% (145) as GST fusions. A *C*-terminal phage display library of 10^9 random peptides was used to screen all the purified constructs. Overall, 50% of the

PDZ domains (82) were analyzed for *C*-terminal ligand consensus sequences. Importantly, the individual PDZ domains were found to be specific and not promiscuous (Figure 10).

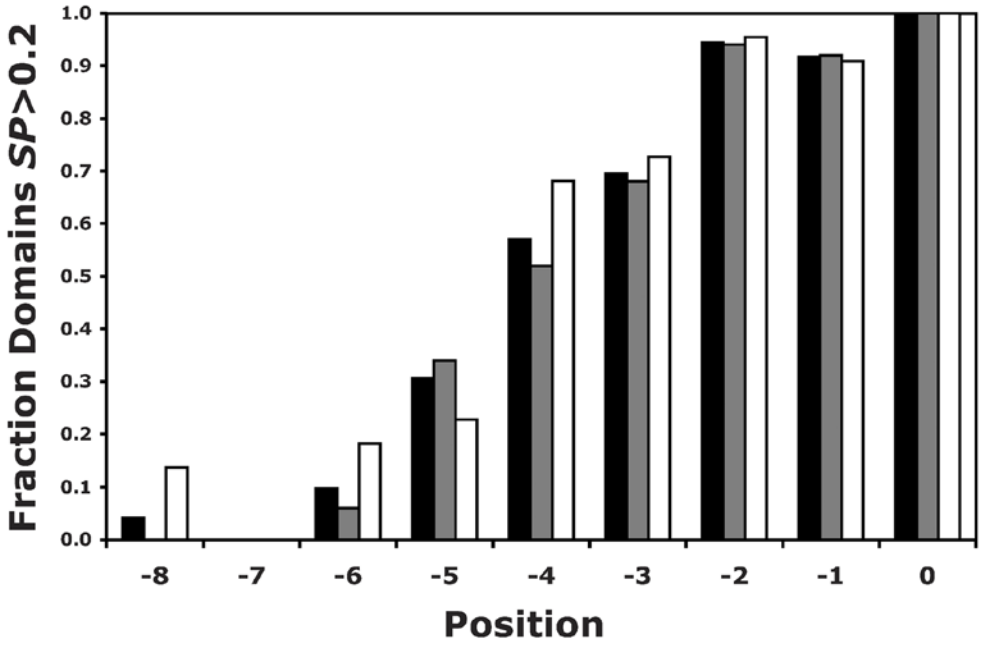


Figure 10. Fraction of PDZ domains exhibiting significant specificity scores. The figure shows specificity scores (SP) greater than 0.2 at each ligand position. Higher SP values correlate to more specificity for a certain amino acid by the PDZ domain. Black bars: all PDZ domains in the study, Grey bars: Human PDZ domains, and White bars: Worm PDZ domains. This figure is reproduced from PLoS Biology [136].

A high correlation between ligand position -2 with αB-1 (the first position in the αB structure) was found, consistent with Bezprozvanny's and Chen's earlier results [131, 135]. A high correlation to ligand position -2 with αB-5 was found as well. It was concluded that ligand position -2 and 0 appears to be directly influenced by proximal PDZ residues; however, the other ligand residues appeared to be influenced by PDZ residues scattered throughout the domain. Significantly, this comprehensive analysis suggests that PDZ domains are robust, remaining active while withstanding substantial mutational pressure [136].

Finally, this study further probed viral genomes and compared them to the ligand consensus sequences they obtained for the human PDZ domains analyzed. Importantly, 89 viral proteins were found that could potentially bind to human PDZ domains, and therefore had the potential to disrupt normal cell processes in human cells. Known viral PDZ binding motifs were confirmed (e.g. human papilloma virus (HPV) and avian influenza [123, 127]) and several new potential interactions involving herpes, vaccinia, myxoma, and fibroma viruses were discovered [136].

Engineering PDZ domains

Three studies on the engineering of PDZ domain ligand binding by laboratory protein evolution and one by rational design have appeared in the literature [63, 64, 70, 137]. The objective of each of these studies was to redesign the targeted PDZ domain's binding interaction with peptide ligand. Table 2 summarizes the protein redesign strategy for each approach.

Table 2: Summary of redesigned PDZ domains. PSD-95/3 is the postsynaptic density 95, third PDZ domain. Af-6 is a RAS associated protein. Omi/HtrA2 is a mitochondrial protein released in the cytoplasm following apoptotic stimuli. NHERF/1 is Na+/H+ exchanger regulatory factor, first PDZ domain. SPR: Ligand binding affinity determined by Surface plasmon resonance; FP: Ligand binding affinity by fluorescence polarization. Qualitative binding analysis refers to in vivo binding analysis using a β-galactosidase reporter system assay. N/A: not applicable, rational protein design was used.

PDZ Template	Library Method	Library size	Screening Method	Binding Analysis	Ref.
Af-6	ePCR	$5.0x10^5$	Yeast two-hybrid	SPR	[63]
PSD95/3	N/A	N/A	N/A	FP	[137]
Omi/HtrA2	ePCR	$7.5x10^5$	Yeast two-hybrid	Qualitative binding analysis	[64]
NHERF/1	ePCR	$1x10^6$	Phage Display	FP	[70]

Schneider *et al.* first attempted to redesign the PDZ domain of the human ALL-1 fusion partner protein from chromosome 6 (AF-6). AF-6 is a component of epithelial cell tight junctions, non-epithelial cell adhesions and interacts with acute lymphoblastic leukemia (ALL-1) protein and Ras protein, both oncogenic proteins [138, 139]. The authors first constructed a PDZ mutant library containing 2 x 10^5 clones, generated by amplifying the PDZ template using ePCR conditions. They then used the yeast two-hybrid screen to select for active binders to four 9mer peptide ligands that wild-type AF-6 PDZ domain did not bind in a preliminary yeast two-hybrid analysis. The library PDZ domain mutants were fused to the activation domain (AD) of GAL4 and the 9mer peptides to GAL4 binding domain (BD). GAL4 is a transcription activator that activates transcription of genes for galactose metabolism when its modular domains, AD and BD, dimerize. Positive mutants were detected using x-gal, a chromogenic substrate of β-galactosidase.

After one round, they isolated one active variant for three of the peptide ligands and two variants for the fourth peptide ligand. Using surface plasmon resonance (SPR), they were able to measure binding affinity for three of the five variants with their target peptide immobilized on the surface. The equilibrium dissociation constants (K_D) ranged from 0.16 to 0.24 μM. However, the binding affinity of the wild-type PDZ domain to the target peptides was not reported and hence the improvement in affinity could not be deduced. The mutations found in the selected variants align with the critical residues in secondary structure elements, αB1 and βB1 and βB5, identified by Bezropanny, Chen and Tonikian as important for determining binding to the (-2) position in the peptide ligand [131, 135, 136].

In 2002, Reina *et al.* rationally designed PDZ domains based on the third PDZ domain of PSD-95 (PSD-95/3), a post-synaptic density protein. Up until that time, it was generally agreed that the peptide ligand position conferring PDZ binding specificity was the (0) and (-2) position. Reina's objectives were two-fold: to demonstrate they could successfully redesign a PDZ domain *in silico* and to redesign the PDZ domain to recognize different peptide ligand residues in positions other than (0) and (-2) with binding dissociation constants (K_D) in the range of the wild-type PDZ domain and ligand, which is 20 ±1.5 μM. They used three different target peptide sequences: two altered ligand positions (-1) and (-3) and the third target altered at positions (-1) and (-4). They chose PSD-95/3 PDZ domain as the model PDZ domain because its binding interaction with its physiological ligand is well characterized and there is a high resolution structure available. They used PERLA (Protein Engineering Rotamer Library Algorithm) [140, 141] to test user-defined mutations in the wild-type complex for stability with the three target peptide ligands. PERLA evaluates the stability of mutations in a given structure and is capable of evaluating multiple mutations at one time. They chose PDZ domain

residues for mutation by visual inspection of the PSD-95/3 complex primarily from the PDZ secondary structure elements, βB, βC and αB. $2.3x10^{9}$, $3.9x10^{22}$ and $1.4x10^{4}$ variant PDZ sequences were evaluated by PERLA for the three target peptide ligands, respectively. One designed PDZ domain for each target peptide combination with the highest PERLA calculated stability score was selected for further experimental analysis.

The two variant PDZ domains designed for binding altered residues in ligand position (-1) and (-3) had dissociation constants (K_D) of 1±0.008 μM and 26 ± 3.2 μM for their target peptide ligands. The variant PDZ domain designed for binding altered residues in ligand position (-1) and (-4) had a dissociation constant (K_D) of 96±11.6 μM and did not meet their objective. The authors further used a yeast two-hybrid analysis to confirm the designed PDZ domains were specific for their target peptide. The designed PDZ domains were specific for their target, but also retained binding affinity for the wild-type peptide albeit at a lower affinity than the parental, unmutated PDZ domain. Reina *et a.l* surmised that peptide positions (0) through (-3) of the peptide contribute to specificity of the ligand binding to the PDZ domain. Peptide position (-4) does not have the same impact since it has fewer interactions with the PDZ domain than the other four positions [137].

Junqueira, *et al*. demonstrated a PDZ domain could be redesigned to specifically target and inactivate an oncogene while probing a PDZ domain's structure through laboratory evolution. They chose the Omi/HtrA2 PDZ domain as their parent template [64]. Omi/HtrA2 is homologous to bacterial proteins DegS and DegP, which displays protease activity at high temperatures and chaperone function at lower temperatures. Omi/HtrA2 is a human mitochondrial protein that is pro-apoptotic when released to the cytoplasm during cellular stress. The physiological Omi/HtrA2 peptide ligand is –LVMI. The authors selected new target peptide ligands recognizing the *C*-terminus of

the human c-Myc protein, an oncoprotein with the sequence –NSCA. c-Myc is critical in cell growth, differentiation and apoptosis [142]. A library with $7.5x10^5$ mutants was constructed using an error prone polymerase, with mutazyme (Stratagene) and screened using the same yeast two-hybrid methodology used previously by Schneider *et al*. for engineering the PDZ domain of AF-6. Junquiera *et al*'s method resulted in selecting fifteen variants, but only one variant was a true positive binder to the novel ligand as determined by a second qualitative β-galactosidase assay. This variant contained two mutations outside the amino acids that make up the ligand binding groove, neither in the PDZ secondary structure elements, αB or βB, nor in the conserved hydrophobic sequence. Expression of the selected, mutant PDZ domain in mammalian cells was shown to target c-Myc, and activate apoptosis effectively, which demonstrated the PDZ domain was successfully engineered for the novel desired target.

Finally, Ferrer *et al.* sought to further optimize an enzyme assay that utilized a PDZ domain as a detection reagent that binds to the new *C*-terminus of an enzyme cleavage product by improving the PDZ domain binding affinity for the cleavage product. In this assay, an HIV protease cleaved a peptide fused to biotin at its *N*-terminus. A purified GST-PDZ domain (glutathione-S-transferase tagged PDZ domain) that binds to the *C*-terminus of the cleavage product was used to detect proteolytic cleavage. The GST-PDZ-cleavage product complex was then incubated with (Eu^{3+})GST-antibody and streptavidin labeled with XL665. The authors detected the peptide product using TR-FRET (time-resolved fluorescence resonance energy transfer) (see Figure 11) [70, 143].

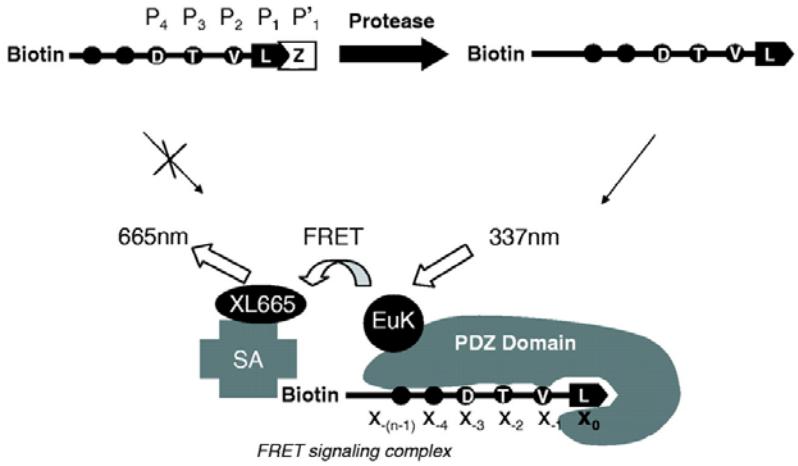


Figure 11. Enzyme Assay using a PDZ domain as the product detection reagent. The substrate is labeled with Biotin at the *N*-terminus. After cleavage, the biotin fused product is bound by the GST-PDZ domain fusion. Anti-GST-Eu^{3+} (EuK in the diagram) and Streptavidin(SA)-XL665 are incubated with the product. The product is excited at 340 nm and the emission is measured at 620 and 665 nm. This figure is reproduced with permission of Oxford University Press [70].

To engineer appropriate PDZ domains for this assay the authors chose two well characterized PDZ domains as the parent templates: the Na^+/H^+ exchanger regulatory factor PDZ1 (NHERF/1) and PSD95/3. Libraries comprising 10^6 mutants with a 1% error rate for each parent template were constructed by error prone PCR. After five rounds of phage panning, the PSD95/3 library did not show any enrichment whereas the NHERF/1 resulted in enrichment of specific binders. Binding affinity was quantitatively measured for the three selected variants by fluorescence polarization using varying concentrations of purified GST-PDZ domains incubated with fluorescent peptide ligand. The binding affinity for the new target peptide, RYLDTVL, was increased by 25- fold while affinity for the wild-type peptide ligand, EVQDTRL was also retained (see Table 3). Two of the three selected variants had a common mutation in position βB:3 which interacts directly with ligand [135, 136]. The consensus mutation for all three selected variants was at βF:3 and was the only mutation that is not near the ligand binding interface. Other mutations were unique to each variant and located in the PDZ

secondary structure element, βC. Importantly, the redesigned PDZ domains improved the sensitivity of the HIV protease assay. The best PDZ domain variant in the HIV protease assay detected cleavage product to 2 nM vs. 10 nM for the wild-type PDZ domain.

Table 3. Binding Dissociation Constants (K_D) for NHERF/1 and selected Variants. G9, H9, and F10 are selected variants of NHERF PDZ1. Data extracted from Ferrer *et al.* [70].

PDZ Domain	Binding Dissociation Constants (K_D) μM	
	RYLDTVL (target)	EVQDTRL (wt)
NHERF/1	16±7	0.31±0.05
G9	0.770±0.100	1.1±0.270
H9	0.620±0.06	0.770±0.17
F10	1.7±0.3	0.410±0.062

ALLOSTERIC PROTEINS

The effect (inhibition or activation) of a ligand binding at a specific site other than the reaction site of a protein is called allostery [22]. Traditionally, this definition applied to oligomeric proteins, but it is now recognized that it can also apply to monomeric proteins [144]. Recently, some have broadened the definition of allostery to include any change to a specific site (a mutation or even pH effect) that results in a change to a distal site [145]. For our purposes, allostery will refer to the effect of ligand binding to a specific site affecting proteolysis at a distal site in the same protein.

Allosteric proteins exist in metabolism, cellular signaling, and gene regulation [146]. The discovery of a protein complex changed by a secondary molecule dates to 1903 with Bohr's study of the hemoglobin protein complex. Another historic example is

Gerty Cori's research confirming that AMP activates glycogen phosphorylation in 1938, part of the research that eventually earned her a Nobel prize [147, 148].

In 1963, Monad named the reversible conformational change that occurs when a ligand binds to a site other than the site where substrate interacts with a protein an allosteric transition [149]. The MWC (Monad, Wyman, Changeux) model describes two potential conformational states: tense (T) and relaxed (R) that are always present in equilibrium with the predominant population determined by ligand binding (see Figure 12) [150]. Later, Koshland, Nemethy and Filmer (KNF) modeled allostery as a sequence of events caused by induced conformational changes (see Figure 13) [151].

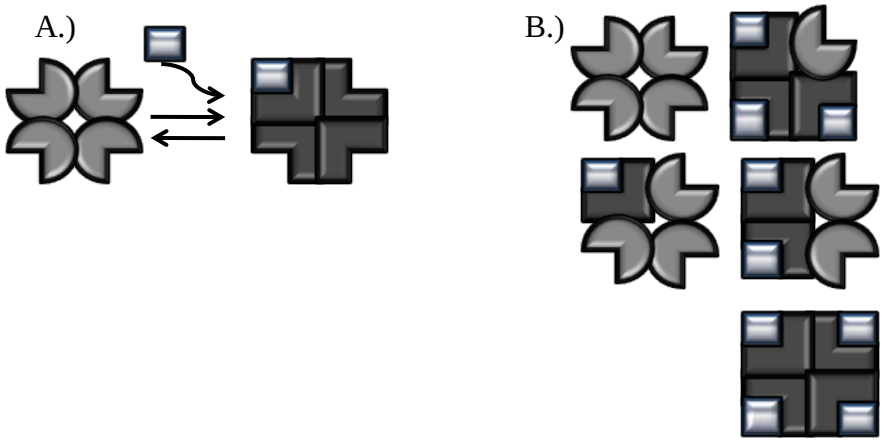


Figure 12. Classic models of allostery.
A.) The representation of the MWC concerted conformation model with the R state (unbound) and T state (bound) in equilibrium. The T state is favored when at least one ligand is bound. B.) The KNF model portrays a sequential series of events with a fixed change resulting from each ligand binding event.

A general molecular mechanism of the allosteric effect is not known. Recently, Ranganathan's group used a computer algorithm called statistical coupling analysis (SCA) to determine which amino acids play an important role in linking two distal events in allosteric proteins [152]. SCA is based on the hypothesis that two linked sites must have evolved together so residues conserved similarly to the residues at the two distal

sites are important to link the two sites. They analyzed three different protein families, a transmembrane signaling receptor family belonging to the GPCRs class, a zymogen serine protease family (chymotrypsin-like), and the classic allosteric model of multi-subunit proteins from the hemoglobin family. Ranganathan *et al.* found that the functional domains appear to be modular and linked by a sparse amino acid network [152]. A second, independent computational approach and set of mutagenesis experiments corroborate the SCA results [153, 154].

Engineering allosteric proteins

The engineering of protein allostery was recently reviewed [155-157]. The main protein engineering strategies for generating allosteric proteins are shown in Figure 13 [156]. In Figure 13(a) an enzyme is split into two fragments and fused to two proteins that associate with each other based on the concentration of a third molecule or ligand. When the ligand is added, the functional protein fragments are brought next to each other and this permits them to form the active conformation.

Strickland *et al.* designed a strategy (Figure 13(b)) that overlaps sequences between two modular, functional protein domains. The two protein domains have complementary terminal α helices: a light activated phototropin (LOV2) domain, which is widely distributed and well characterized modular signaling domain, and a DNA binding domain from the bacterial *trp* repressor (TrpR). The fused protein includes the two functional domains and a shared, single α helix. The α helix is preferentially bound to the LOV2 domain, until LOV2 is photoexcited. Light is the allosteric effector that alters LOV2's conformation, causing the α helix to preferentially bind to the TrpR domain, which disrupts TrpR from DNA binding and permits transcription activation of desired genes [28].

Circular permutation of two proteins has been used to engineer allosteric protein switches as shown in Figure 13(c). Circular permutation is effectively linking the ends of gene sequences and then linearizing the sequence by cleaving at different nucleotide positions resulting in different terminal sequences. The positions of the functional domains of separate proteins are varied using circular permutation until they influence each other. This strategy requires an iterative approach to find the sequence that corresponds to two active and linked functional domains.

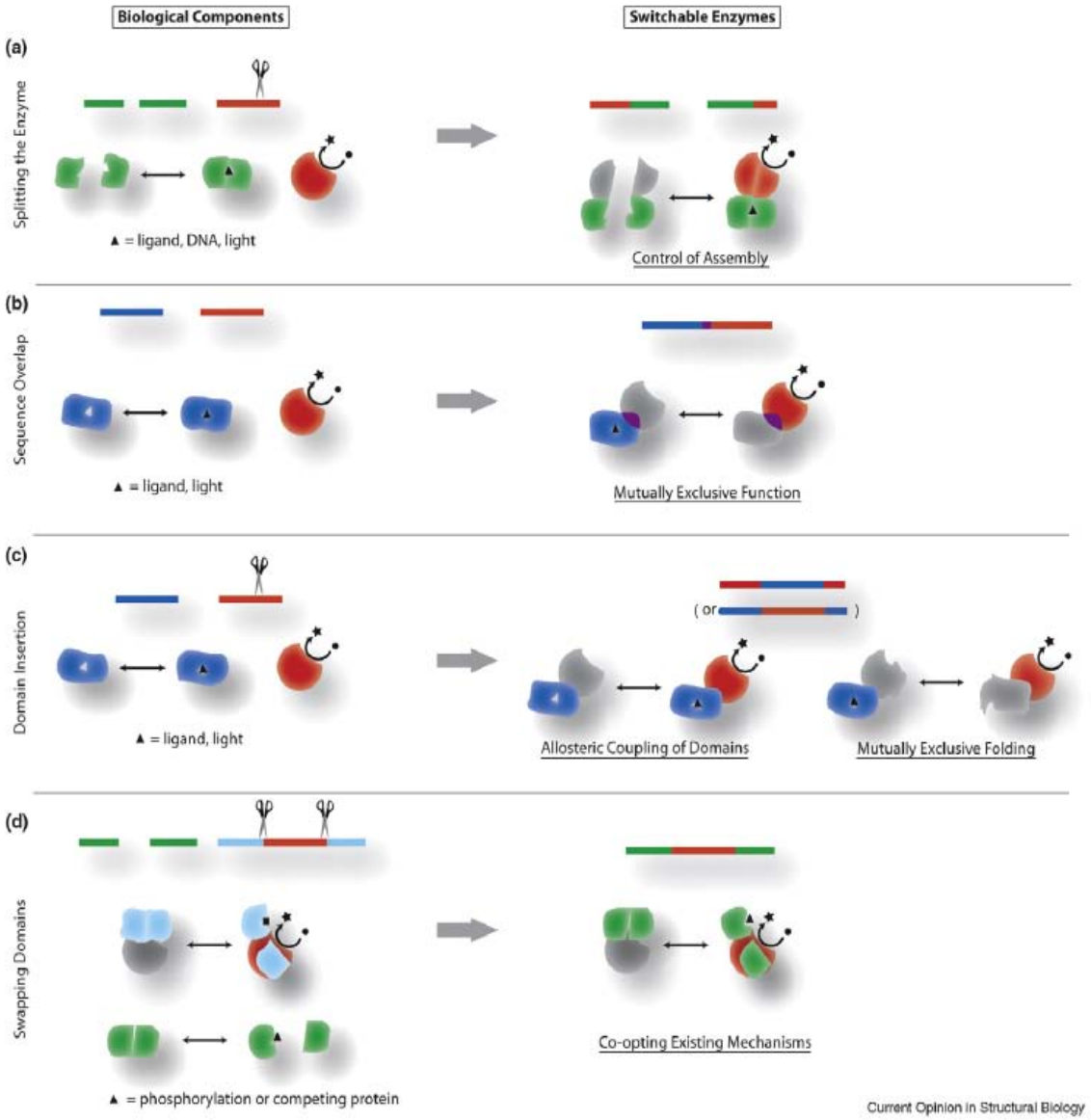


Figure 13. Strategies for designing allosteric enzymes.
The biological components are the wild-type proteins from which the "Switchable Enzymes" were derived. Ostermeier named each switching mechanism. The names are underlined text. DNA sequences are depicted as lines and the proteins are represented as geometric shapes. The enzyme domain is in red. The color gray indicates the enzyme domain is less active. The effector ligand or signal is a black triangle. Note that (b) has not actually been demonstrated with enzyme activity but with binding affinity (binding occurred when the domain was "activated"). This figure is reproduced with permission from Elsevier [156].

Finally, the fourth strategy for generating new allosteric proteins is to extract the domains from existing allosteric enzymes and fuse them to another allosteric enzyme to change the activation signal or the protein's activity (Figure 13(d)). Several successful applications of this strategy have been reported and this strategy has been applied with the most success [25, 29, 158, 159]. For example, a guanine nucleotide exchange factor (GEF) was engineered to activate with a novel effector by fusing it with another protein's PDZ binding domain [29]. GEFs regulate GTPases, which in turn regulate the actin cytoskeleton and are responsible for morphological changes on the cell surface. Yeh *et al.* fused the GEF active site between a syntrophin PDZ domain and a short peptide ligand sequence that the syntrophin PDZ domain recognized and that could also be phosphorylated by protein kinase A (PKA) (see Figure 14). The GEF was inactive as long as the PDZ domain bound the fused peptide ligand. GEF was activated by adding PKA, which phosphorylated the peptide sequence and disrupted the PDZ domain-peptide ligand complex. Engineered allosteric GEFs were cloned and analyzed for two different GEFs in rat cells. The engineered allosteric GEFs were successfully activated by PKA *in vivo* as demonstrated by the appropriate change in cell surface morphology.

Figure 14. A specific example of "Co-opting Existing Mechanism" allosteric engineering strategy.
The functional domain of a GEF (the light grey rectangle) is fused to the syntrophin PDZ domain and a short peptide (the dark grey diamond) that is recognized by the PDZ domain and protein kinase A (PKA). When PKA is added, it phosphorylates the peptide sequence, dissociating the PDZ domain and creating the active structure of the GEF functional domain. Figure constructed with generalized nomenclature based on system designed by Yeh et al.[29]

All these strategies for redesigning an allosteric protein use iterative rounds of rational protein design followed by the cloning, expression and testing of variants. Laboratory evolution has not yet been applied to redesign allosteric proteins or modules.

DegS

In this research, laboratory protein evolution was used to engineer the ligand binding specificity and affinity of the DegS PDZ domain separate from the DegS protease domain. The effects of altering the ligand binding properties of the PDZ domain on the allosteric activation of DegS were then analyzed after fusing the engineered PDZ domain to the wild-type DegS protease domain. Before discussing the results in the following two chapters, background information on the allosteric protein template, DegS, follows.

DegS and the *E. coli* σ^E Unfolded Protein Response In *E.coli*, as in other organisms, RNA polymerase (RNAP) is a multi-subunit protein complex that initiates transcription of genes when bound to a σ factor creating the active RNA holoenzyme (Eσ). The RNA holoenzyme complex transcribes genes based on the recognition of promoter sequences by the σ factor. The predominant σ factor bound to the RNAP complex will change depending on the growth phase or environmental conditions, which cause varying cellular concentrations of the σ factors.

There are at least seven different σ factors in *E. coli* and over 350 different transcription activator or repressor molecules that combine to make up the RNAP complex, which makes a complex transcription regulation scheme that I will simplify by describing only an example of the σ factor effect [160, 161]. In *E. coli*, the general housekeeping σ factor, σ^{70}, is bound to RNAP during normal environmental conditions. However, during high temperature conditions, cellular proteins denature, which can lead to protein aggregation and potentially cell death. Thus, in response to high temperatures,

the alternative σ factor, σ^{32}, binds to the RNAP to transcribe additional folding chaperones and proteases to refold and degrade denatured proteins and anabling the cell to survive.

In 1989, Erikson and Gross discovered a distinct *E. coli* σ factor that initiated transcription under conditions of outer membrane stress, the σ^E factor [162]. The σ^E factor was one of the founding members of the extracytoplasmic function (ECF) σ factors, responsible for maintaining the integrity of the outer cell membrane[163]. Consequences for complete inhibition and over-expression of the σ^E factor are membrane blebbing and eventual cell lysis, which indicates the regulation of the σ^E factor is critical and it is essential for cell survival [164-166]. There appears to be potentially two pathways that induce transcription of genes by $E\sigma^E$ [165, 167]. The DegS pathway is the primary regulatory mechanism and the only one elucidated at this time.

DegS is a trimeric periplasmic anchored serine protease essential to *E. coli* due to its role in σ^E activation and regulation [168, 169]. The σ^E factor is tethered to the cytosolic side of the periplasmic membrane by the transmembrane protein RseA (**R**egulates **s**igma **e** **A**). This interaction is very stable with an equilibrium dissociation constant (K_D) that is less than 10 pM [170]. RseA must be degraded to disrupt the binding between the σ^E factor and RseA. Unfolded OMPs (outer membrane porins) in the periplasm initiate a three step proteolytic cascade of RseA (Figure 15) [167, 168, 171-175]. The OMPs bind to at least one of the DegS PDZ domain subunits and allosterically activate the DegS protease domain. The bulky, periplasmic region of RseA is released to the periplasm after it is cleaved by DegS between RseA residues: Val148 and Ser149. This cleavage removes the bulky periplasmic region of RseA, which disrupts the association between RseA and RseB. RseB inhibits proteolysis of RseA by RseP. Once RseB is dissociated, RseP, a periplasmic anchored metalloprotease, is able to cleave

RseA in a transmembrane region near the cytoplasmic side of the periplasmic membrane. The intra-membrane proteolysis of RseA by RseP releases RseA with the bound σ^E factor into the cytoplasm. In the cytoplasm, RseA is a ubiquitous substrate for cytosolic proteases, including ClpAP, ClpXP, Lon, and FtsH. Degradation of RseA releases the σ^E factor and permits transcription of genes including transcription factors and regulatory genes (*rpoH, rpoE, rpoD, rseA/B/C*), periplasmic folding chaperone genes (*skp, dsbC, fkpA, surA*), protease genes (*depP, rseP, clpXP, lon*), genes involved in lipopolysacharide (LPS) biogenesis (*htrM, lpxD*), and other genes with unknown functions [176, 177].

Chaba *et al.* evaluated the kinetics of the DegS proteolysis cascade to determine the important rate limiting step using pulse-chase experiments with L-[^{35}S]-methionine pulse-labeled cells. The half life of full length RseA under steady state cellular conditions is 8 minutes. In envelope stress conditions, the full length RseA half-life is reduced to 1 minute. The half life of the DegS cleaved RseA product located in the periplasmic membrane is less than 20 seconds. The half life of the RseP cleaved RseA cytosolic fragment, is less than 30 seconds. The data reveals the activation of the σ^E factor corresponds to the activation of DegS proteolytic activity, the rate limiting step [170].

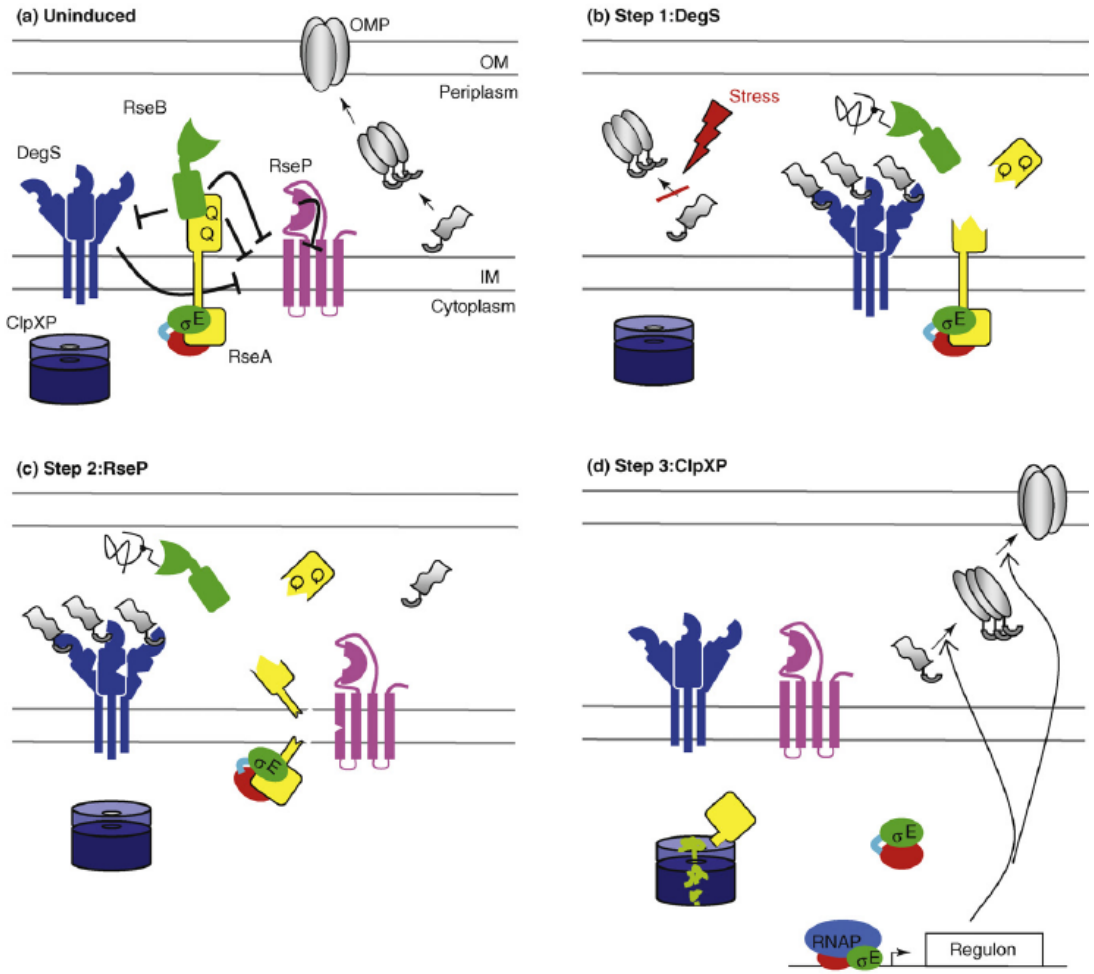


Figure 15. Unfolded OMPs in the periplasm initiate a three step proteolytic cascade of RseA .

(a) When the pathway is in the uninduced state, DegS (blue) is inactive and RseB (green) is bound to the periplasmic domain of RseA. DegS, RseB, the glutamine-rich regions of RseA (indicated by Q's), and the PDZ domain of RseP inhibit RseP (magenta). sE (red and green) is bound to RseAcyto. Outer membrane porins (OMPs) (gray-shaded shapes) enter the periplasm via the Sec secretion machinery (not shown) and are escorted to the outer membrane by a series of chaperones. (b) Upon cell envelope stress, OMP folding is disrupted. The *C*-terminal peptides of unfolded OMPs bind to DegS, activating DegS. RseB is also removed from RseA, perhaps by lipoproteins as diagrammed. DegS cleaves RseA in the periplasmic domain. (c) RseP cleaves RseA in the transmembrane region. (d) ClpXP (light and dark blue) degrades RseAcyto releasing sE to interact with RNA polymerase (RNAP) and direct transcription. sE regulon members restore proper OMP folding, resetting the pathway. OM, outer membrane. IM, inner membrane. The figure is reproduced with permission from Elsevier [178]

The Mechanism of DegS Activation Since 1999, the DegS activation mechanism has been investigated [166, 168, 170-173, 179, 180]. It is known that DegS is allosterically activated by its PDZ domain binding to the *C*-terminus of the *E. coli* OMP consensus sequence, Y-X-F (a Class II PDZ binding motif, Φ-X-Φ) (see Figure 16) [172, 181]. There are two hypotheses proposed that describe the molecular mechanism of DegS activation: "The Peptide Activation Model" proposed by Clausen's group and the "Inhibition Relief Model" proposed by Sauer's group [181-185]. Common ground between the two hypotheses is agreement that ligand binding leads to activation and different ligands result in different DegS proteolysis rates [181, 183].

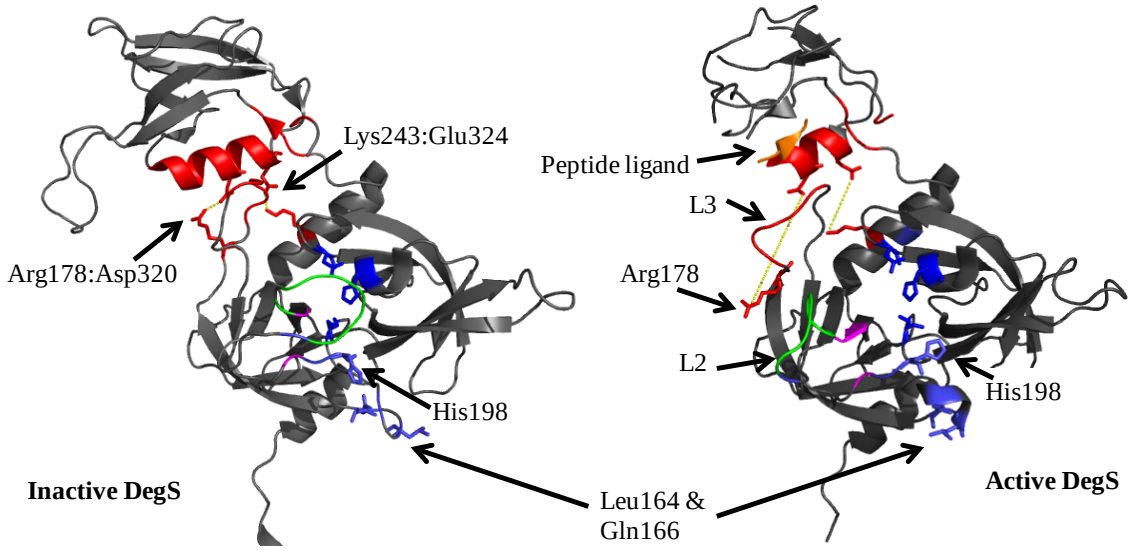


Figure 16. Ribbon presentation of one subunit of the DegS trimer.
Loop 3 (L3) (red) mediates communication between the PDZ and protease domains. Two salt bridges (Arg178:Asp320 and Lys243:Glu324) tether L3 to the PDZ domain. When the PDZ domain binds ligand (shown as a peptide and in orange), the salt bridges are disrupted and L3 conformation changes and impacts L2 and other key residues (represented in stick format and in blue) and allows formation of a functional catalytic triad comprising residues His96, Asp126, and Ser201, which are shown in stick mode and darker blue. These were built using PyMol and data from the solved structures: Inactive DegS (PDB:1TEO) and Active DegS (PDB:1SOZ)[104, 184].

Clausen's group proposed the "Peptide Activation Model" to explain the allosteric activation of DegS. Clausen's group hypothesizes that peptide ligands differentially activate the protease domain by binding to the PDZ and protease domain, which relieves PDZ domain inhibition of the DegS protease domain and the peptide ligand directly activates the DegS protease domain by interacting with protease domain Loop 3 through the (-1) peptide ligand residue (see Figure 17) [181]. Clausen's group analyzed binding affinity between the PDZ domain and peptide ligand by isothermal calorimetry (ITC) and measured relative DegS activity using SDS-PAGE to evaluate the DegS reaction products after incubation with different peptide ligands. Their data revealed peptide ligands with higher binding affinity for the DegS PDZ domain also had higher DegS proteolytic activity. Clausen's group's crystal structure data indicates that DegS proteolysis differences may be explained by the different positions of Loop 3 resulting from interactions with the (-1) peptide ligand residue, although they did not propose the molecular mechanism of how Loop 3 changed protease activity and whether it was the initial protease activation rate or maximal protease activity that was impacted by Loop 3.

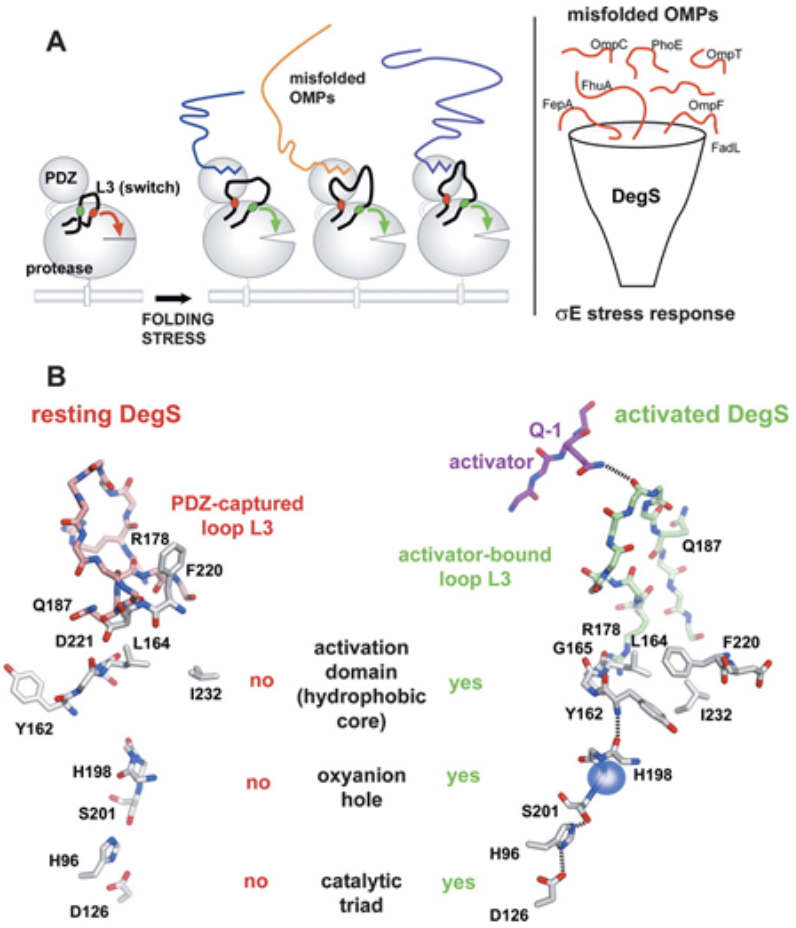


Figure 17. Peptide Activation Model of DegS (A) The left panel illustrates activation of DegS by different OMP *C*-termini. Loop L3 is highlighted with its inhibitory (red) and activating (green) structural elements. Molecular details of both inhibitory and activating processes are given in B. In latent DegS, loop L3 directly inhibits protease function by disrupting the activation domain. Binding of the allosteric activator to the PDZ domain triggers a switch of this loop into its active position, where it now supports the setup of a functional proteolytic site. The right panel illustrates the cellular function of DegS acting as a mechanistic funnel to integrate the information from different mislocalized OMPs into the σE stress response. (B) Working model for how DegS switches from the resting to the activated state. Key residues that are important for regulation and for signal propagation are labeled. Loop L3 of the resting DegS is drawn in red, loop L3 of the active DegS is in green. Figure and caption reproduced to illustrate the "Peptide Activation Model" [181].

Sauer's hypothesis, "Inhibition Relief Model of DegS", uses the classic MWC allosteric model as its basis to explain the allosteric mechanism of DegS activation [149, 182, 186]. Sauer's group hypothesizes DegS exists in equilibrium of tense (inactive) and

relaxed (active) states that shifts upon ligand binding and ligand binding is cooperative between the three DegS subunits. The influence of RseA substrate binding to cause the concerted shift from tense to relaxed state is less because RseA binding affinity to the tense (inactive) state is low reducing the potential cooperative effect of substrate binding. The stronger ligand binders shift equilibrium faster than weaker ligand binders resulting in different DegS protease activation. Their results show that different OMP peptide ligand sequences result in different maximal activation of DegS activity as well as different rates of activation. They found no difference when they varied the (-1) ligand peptide residue in the maximal activity of DegS but they did find a difference in the activation rate of DegS between these ligands. They found a 35-fold maximal DegS activity difference when they changed residues upstream of the (-3) peptide ligand residue.

Sauer's group fitted their experimental data to equations describing the MWC allostery model using iterative subroutines in ORIGIN (global fit; OriginLab) and MATLAB (LSQNONLIN (least squares non-linear fit) algorithm: MathWorks) to demonstrate their data fitted the MWC model and could explain the low level DegS basal activity [183]. They calculated that without OMP peptide ligand and RseA substrate, the ratio of inactive to active DegS is 15,000:1. They added saturating amounts of RseA (200μM) and no activating ligand and calculated the ratio of inactive to active DegS to be 400:1. With saturating amounts of the best wild-type activating peptide ligand (DNRDGNVYYF, 30 μM) and no RseA, they calculated the ratio of inactive to active DegS to be 8:1. Saturating amounts of ligand and substrate changed the ratio of inactive to active DegS to 1:5. Their hypothesis explains the sensitivity of DegS in the presence of substrate and ligand and the low basal rate of DegS activity that is essential for cell viability since active DegS is present even without activating peptide ligand present.

Their biochemical and structure results indicate the (-1) peptide ligand residue is not important to activate DegS as the "Peptide Activation Model of DegS" hypothesis suggests; however, their data does not eliminate the possibility of the (-1) ligand position playing a role in DegS activation since structural data of Loop 3 is not conclusive.

Both groups' biochemical and structural data do not disprove either hypothesis. The structures of ligand bound PDZ domains have low resolution with 40 to 70% of the residues defined. Protease domain Loop 3 structural elements vary between the subunits of the DegS oligomers crystallized making it difficult to conclude if the peptide ligand is causing the Loop 3 movement [34, 181, 183, 184, 187]. Understanding the molecular mechanism of DegS activation would assist in redesigning DegS. At this point, redesigning DegS could contribute to a better understanding of the molecular mechanism.

For this dissertation, I changed the (-2) peptide ligand residue of the binding peptide ligand to alter the DegS PDZ domain binding affinity from the conventional Class II PDZ domain (Φ-X-Φ) to the Class III PDZ domain (E/D-X-Φ) to avoid potentially impacting allostery by changing the (-1) ligand position. Clausen had shown that an Asp (D) in the (-2) ligand position neither activated DegS nor had detectable binding to the DegS PDZ domain [181]. I used laboratory protein evolution to redesign the DegS PDZ domain for a novel ligand and to increase binding affinity towards the wild-type ligand and that is the subject of Chapter 2. Chapter 3 includes an analysis of the impact on allostery of the engineered PDZ domain variants fused to the wild-type DegS proteolytic domain.

CHAPTER 2: ENGINEERING THE SPECIFICITY OF A PDZ DOMAIN TO BE USED FOR PEPTIDE DETECTION

CHAPTER SUMMARY

Objective: Develop a method to rapidly screen library protein mutants of the DegS PDZ domain to meet two objectives. The first objective was to increase the PDZ domain's binding affinity towards wild-type peptide ligand. The second objective was to create binding affinity toward a novel peptide ligand, not recognized by the DegS PDZ domain.

Approach: The PDZ domain from the *E. coli* DegS protein was fused to an *N*-terminal six residue sequence derived from *E. coli* lipoprotein NlpA [86], anchoring the PDZ domain to the periplasmic membrane. The outer membrane was then permeabilized permitting a 1 kDa peptide ligand conjugated to a fluorophore to access the periplasm. The binding of labeled peptide to the PDZ domain renders the cells fluorescent and thus permits their selection by flow cytometry. Selected cells were subjected to iterative rounds of growth and flow cytometric sorting with decreasing ligand concentration.

Results: The DegS PDZ domain was successfully engineered for higher affinity (i.e. decreased equilibrium dissociation constant, K_D) by 20-fold, from 1.3 ± 0.2 μM to 60 ± 20 nM. Additionally, even though the DegS PDZ domain does not exhibit detectable binding to Class III peptide ligands with a *C*-terminal amino acid sequence of Asp-Tyr-Phe (D-Y-F), random mutagenesis and screening led to the isolation of a variant with a K_D of 770 ± 170 nM.

INTRODUCTION

PDZ protein binding domains interact with protein ligands with equilibrium binding dissociation constants (K_D) in the μM range. Protein interactions in the μM range are reversible yet strong enough to be sensitive to cellular protein ligand concentrations, which is typical of proteins within signaling pathways or involved in assembling transient molecular complexes. In contrast, interactions that have stronger

equilibrium binding dissociation constants, in the nM or pM range, are typically irreversible in the cellular environment and require proteolysis to dissociate the protein interaction [188]. This investigation's objective is to use laboratory protein evolution to increase a PDZ domain's binding affinity from the μM to the nM range for its wild-type peptide ligand and to change its specificity to a novel peptide ligand.

Throughout this chapter, the nomenclature used to reference amino acid positions within the PDZ domain is standardized using PDZ secondary structure elements as modeled in Figure 1 with the third PDZ domain of PSD-95 (postsynaptic density protein 95) [189]. The conventional PDZ binding classification of classes I through III that are characterized by the three *C*-terminal amino acids of the peptide ligand and established by Songyang *et al.* and Stricker *et al.* in 1997; will be used [102, 130]. We chose the target novel peptide ligand by changing ligand position (-2), which is recognized as an important peptide ligand residue in determining PDZ domain specificity as demonstrated by the conventional PDZ classification system based on the (-2) ligand residue (see Table 1) [74, 102, 130, 135].

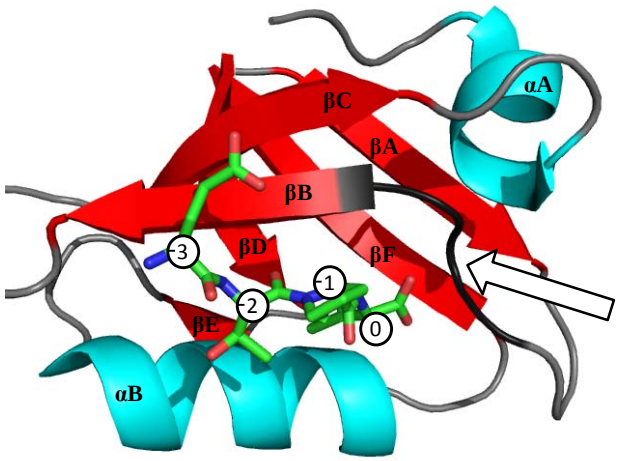


Figure 1. Structure of PSD-95/3 bound to ligand peptide illustrating PDZ domain nomenclature used in this dissertation (PDB:1TP3 [103]). The α-helix structures are in blue and labeled alphabetically from N- to C-terminus. The β-strand structures are in red and also labeled alphabetically from N- to C-terminus. The peptide ligand is labeled with "0" indicating the C-terminal residue. The conserved hydrophobic ligand binding pocket, represented by GLGF in this protein, is highlighted in black in the loop between βA and βB and indicated with a white block arrow. This diagram was generated using PyMol [104].

Table 1: Conventional PDZ Classification by the characteristics of the peptide ligand bound. X: any amino acid, Φ: any hydrophobic amino acid, Ψ: any aromatic amino acid, *: *C*-terminus.

PDZ class	Peptide Ligand Sequence (P_{-2} –P_{-1} –P_0)
Class I	S/T-X- Φ*
Class II	Φ/Ψ-X- Φ*
Class III	E/D-X- Φ*

Four previous investigations have engineered a PDZ domain to bind to novel peptide ligands and are described in detail in Chapter 1. Two of these four previous investigations engineered the PDZ domain to bind to a novel peptide ligand using laboratory evolution and quantified the novel binding affinity. Using a combinatorial library generated by error prone PCR and screened by a yeast two-hybrid screen, Schneider *et al.* redesigned the AF-6 PDZ domain (human ALL-1 fusion partner protein

from chromosome 6 that is found in epithelial tight junction protein complexes) to recognize three novel peptide ligands with varied residues between ligand positions (-6) to (0) with K_D values between 0.12 and .24 μM measured by surface plasmon resonance. The wild-type AF-6 PDZ domain did not have detectable binding to the novel peptide ligands according to the yeast-two hybrid screen[63]. Ferrer *et al.* redesigned the Na^+/H^+ exchanger regulatory factor PDZ1 (NHERF/1) to recognize a peptide ligand that differed from the wild-type peptide ligand in positions: (-7 → -5) and (-1) with a combinatorial library generated by error prone PCR and screened by phage display. The wild-type NHERF/1 PDZ domain equilibrium binding dissociation constant (K_D) of 16 μM was improved to a K_D of 660 nM as measured by fluorescence polarization using GST-PDZ fusions [70].

We chose the *E. coli* DegS PDZ domain for the parental PDZ template, which has no detectable binding to the chosen novel peptide ligand target as measured by fluorescence polarization. Furthermore, the long-term objective of this study is engineering an allosteric cascade wherein an engineered PDZ domain binds a peptide ligand, leading to the allosteric activation of a protease domain. One approach to engineering such an allosteric protease is to engineer the domains as separate entities. The DegS PDZ binding domain has well-characterized allosteric activation when expressed with its corresponding proteolytic domain [172, 182-186].

Our lab has pioneered several *E. coli* cell-based methods for the screening of combinatorial libraries of single chain antibody (scAb) fragments for the isolation of variants with improved binding affinity and expression. [5, 82, 85, 86]. Flow cytometry with FACS is used to sort the combinatorial protein libraries for variants that bind to the target fluorescent peptide ligand or antigen. These methods were the basis for the

methodology developed in this research to screen the DegS PDZ domain for improved binding to the wild-type peptide ligand and a novel peptide ligand.

MATERIALS AND METHODS

Fluorescent peptide ligand Synthesis

10-14 mg of each peptide were purchased from GenScript USA Inc. (Piscataway, NJ) at >98% purity. BODIPY®-FL-SE (4,4-difluoro-5,7-dimethyl-4-bora-3a,4a-diaza-*s*-indacene-3-propionic acid, succinimidyl ester) was purchased from Molecular Probes (Eugene, OR). For substrate synthesis, peptides were dissolved to make a solution of 10 mg/ml in 70-90% water and 10-30% DMF. DMF volume was varied dependent on peptide solubility. The final reaction contained 2.5 mM peptide, 1mM BODIPY®-FL-SE, 25 mM DMAP (4-dimethylaminopyridine), and 70% DMF. The reaction components were mixed well and incubated at room temperature overnight in the dark. The reaction mixture was quenched with 4 volumes of 0.3% TFA the next day after verifying the reaction was complete by HPLC. The product was purified by FPLC using reverse phase column with a 10% to 50% acetonitrile gradient and 0.1% TFA. The purified product was lyophilized and resuspended in buffer (pH = 7.4) to provide a 1mM solution of purified peptide that was confirmed to be the desired product by ESI-MS. Yield varied between 50 and 60%.

Library Construction

Random Library Using Error-Prone PCR A pool of random mutants was constructed using an error-prone polymerase chain reaction (PCR) with varying balance of nucleotides and $MnCl_2$ to obtain a nucleotide error rate of approximately 0.8 to 1.0% [190]. The following primers were used to generate the error prone PDZ domain: 5’GATGGTCGCGTGATCCGCGGCTACATTGGTATCGGCGGACG and

5’TTTTAAGCTTTCCGCGGTTAATTGGTTGCCGG. The ePCR product was digested with the Sfi (NEB) restriction enzyme and ligated into the pAPEX vector with the chloroamphenical resistance gene using T4 ligase (NEB) and standard molecular biology protocols. The desalted ligation reaction mixture was then transformed into electrocompetent *E. coli* Jude 1 cells [(DH10B F’:: Tn10(Tet^r)] cells and recovered on selective media. Aliquots of the pooled cell culture was sub-cultured into selective liquid media to an OD_{600} of 0.2 and grown to OD_{600} of 2.0. Some of this culture was aliquoted into 10% glycerol solution and stored at -80 °C. Some of this culture was lysed and the DNA isolated and then stored at -20 °C for future use.

Libraries of mutants with targeted mutations Libraries of mutants with specific nucleotide sites mutated were made with standard PCR protocols using oligonucleotides containing specific NNS codons. The PCR product was digested with the Sfi (NEB) restriction enzyme and ligated into the pAPEX vector with the chloroamphenical resistance gene using T4 ligase (NEB) and standard molecular biology protocols. From this point, the library construction protocol was the same for the ePCR products as described above.

Flow Cytometry Analysis

E.coli Jude1 [(DH10B F’:: Tn10(Tet^r)] was transformed with pAPEX encoding the controls, empty vector, and the library. Controls included the wild-type DegS PDZ domain, an unrelated protein, and an inactive DegS PDZ domain. The inactive DegS PDZ domain was constructed by replacing the four residues in the conserved region that binds to the *C*-terminus of the peptide ligand to alanines. Cell cultures of the transformation products were grown to saturation overnight with the appropriate antibiotic and 2% glucose. These cultures were then sub-cultured at a dilution of 1: 100 in fresh media with 2% glucose and appropriate antibiotic at 30°C. The cultures were

grown to an OD_{600} of 0.5 to 0.8 (approximately 2-3 hours) and then induced with 0.1 mM IPTG (isopropyl β-D-1-thiogalactopyranoside, Sigma-Aldrich) for 3 to 4 hours at 25 °C. The cultures were normalized to an OD_{600} of 5 units, and centrifuged to remove the media. The cells were then rinsed gently with PBS (phosphate buffered saline) once, centrifuged to remove the PBS rinse and then stored as a cell pellet overnight at 4 °C. The next day, the cultures were suspended in a permeabilization solution of 50 mM Tris-HCl and 150 mM KCl with the pH adjusted to 7.4. 49 µl of the cells were incubated with 1 µl of labeled peptide at the appropriate dilution to provide final concentrations of 10 to 0.1 µM of fluorescently labeled peptide ligand (BODIPY®-FL-peptide). The mixture was incubated for 30 minutes at 4 °C. After 30 minutes, the reaction mixture was centrifuged and the supernatant carefully removed. The labeled cells were next suspended in 50 µl of the permeabilization buffer. Approximately 5 to 10 µl of this reaction mixture were added to 1 ml of sheath fluid (PBS), depending on the cell flow rate measured by the Becton Dickinson FACSAria instrument. The target cell flow rate achieved 1,000 to 3,000 events per second. Cells were monitored and analyzed using the Becton-Dickinson FACSAria instrument with the following settings: side scatter (cell cytosolic attributes) was set to 250 nm and forward scatter (cell size and shape attributes) set to 225 nm. The FITC laser was set to detect 500 nm. The threshold was set with forward scatter at 200 nm and the side scatter at 1000 nm.

Library Sorting

Library sorting was performed using a Becton Dickinson FACSAria instrument, with a sorting gate based upon the control and library fluorescence units measured that day to select the 1 to 10% of the highly fluorescing library cell members. A total of ~$2x10^7$ cells were sorted in ~90 minutes, collecting approximately 2 $x10^5$ cells. The collected cells were grown on agar media with the appropriate antiobiotics (34 µg/ml of

chloramphenicol for pAPEX) and supplemented with 2% glucose overnight at 30°C. After 12-14 hours, an average of 80% of the collected colonies grew. The selected cells were then collected and pooled and their density measured. These cultures were sub-cultured at a dilution of 1: 100 in fresh media with 2% glucose and appropriate antibiotic at 30°C. Aliquots were frozen for reference or for repeating a round, if necessary. Again, the cultures were grown to an OD_{600} of 0.5 to 0.8 (approximately 2-3 hours) and then induced with 0.1 mM IPTG) for 3 to 4 hours at 25 °C. The cultures were normalized to an OD_{600} of 5 units, and centrifuged to remove the media, following the flow cytometric protocol as described previously. Each cycle represented one round. Each library was sorted four to seven rounds. After round three or four, selected colonies were randomly sequenced. Sorting was stopped when: 1) libraries converged on one or two sequences or 2) consensus sequences were apparent and (or) 3) enrichment (increase of fluorescent signal between rounds) did not progress for three sequential rounds.

Purification of PDZ Domain

All the PDZ domains were cloned into pGex6p vectors using a ligation independent cloning method [74] constructing a PDZ domain fused to a GST protein. The sequences were confirmed after transformation and DNA plasmid purification. The sequenced PDZ-GST vectors were transformed into BL21(DE3) cells for expression and purification. Saturated cultures were grown overnight from a single colony and then sub-cultured 1:100 dilution in 2xyt media supplemented with ampicillin and 2%glucose. The sub-culture was grown to an OD_{600} of 0.5, induced with 0.1 mM of IPTG, and grown for an additional 3 to 4 hours. The cell pellets were collected by centrifugation and stored at -20 ºC for one to four days, until purification. The cell pellet was resuspended in 0.1 culture volume of Lysis Buffer (chilled PBS) and the cells lysed using a French pressure cell. The resulting lysate was centrifuged at 16,000 rpm, in rotor JA-20, at 4°C. The

soluble fraction was applied to 5% lysis volume of gluthathione sepharose beads packed in a 5 ml disposable column and pre-equilibrated with Lysis buffer. The soluble fraction was passed through the column via gravity flow. The column was then washed with two column volumes of Wash Buffer (same as Lysis buffer) followed by two volumes of Protease Buffer (50 mM Tris-HCl, 150 mM NaCl, pH 7.5) via gravity flow. HRV 3C protease and buffer was prepared in accordance to the manufacturer (Accelagen, San Diego, CA). The protease was added to the column and incubated for 16-20 hours at 4 °C and cleaved the PDZ domain from the GST tag anchored to the gluthathione sepharose beads. The collected product was then concentrated to 100 to 200 uM. Purity was confirmed by 16% SDS-PAGE gel visualized with gelCodeBlue reagent (Thermo Scientific, Rockford, IL). Concentration was determined with a BCA assay (Thermo Scientific, Rockford, IL) using BSA as a standard and analyzing all samples in duplicate.

Fluorescence Polarization Analysis

Fluorescence anisotropy was used to determine milli-polarization (mP) level for a range of eight protein concentrations with a set concentration of fluorescent peptide ligand (20 or 80 nM) in 20µl volume using a NUNC 384 well plate and read by the PerkinElmer Wallec EnVision instrument. Each point was measured in triplicate. The mP level is obtained from the following equation:

$$mP_L = 1000*(S-G*P)/(S+G*P)$$

S and P are background fluorescence polarization readings using emission filters parallel and perpendicular (respectively) to the excitation filter. The G factor is a grating factor that is instrument and assay dependent. The mP values were then used in the following equation to determine the K_D by non-linear regression analysis using Kaleidograph:

$$mP_{Lmax}[PDZ]/K_D+[PDZ] = mP_L$$

RESULTS

Flow Cytometric Screening Strategy

The wild-type selection peptide, YYF, (see Figure 2) is a ten amino acid peptide that has been demonstrated to bind to the wild-type PDZ domain of DegS [172, 182]. YYF was used to screen variants of the DegS PDZ domain for increased affinity towards a wild-type preferred peptide ligand. It is a Class II PDZ peptide ligand with an aromatic residue (Tyr) in the (-2) position. The novel selection peptide, DYF, is a similar ten amino acid peptide except for a negatively-charged, acidic amino acid (Asp) in the (-2) position, thereby creating a Class III PDZ peptide ligand. DYF was used to meet the second objective: screening for variants with an affinity toward a novel peptide ligand. Both peptides were conjugated to the BODIPY® (4,4-difluoro-5,7-dimethyl-4-bora-3a,4a-diaza-*s*-indacene-3-propionic acid) fluorophore.

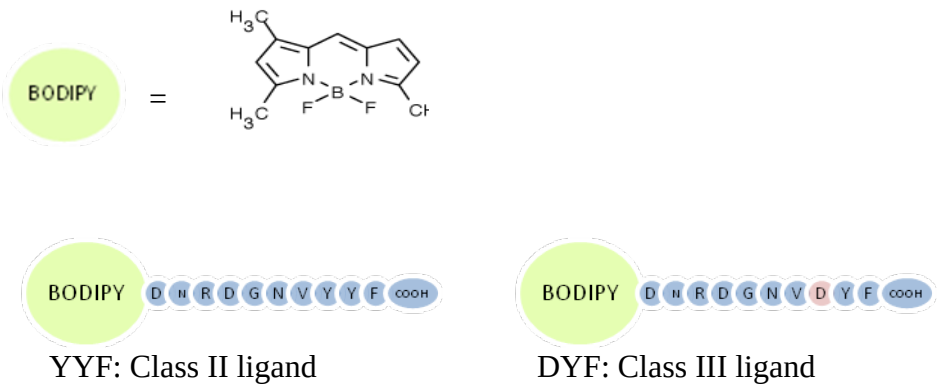


Figure 2. Fluorescent peptides synthesized and used for FACS to select PDZ domain variants. The same fluorescent ligands were also used for fluorescence polarization analysis to quantify the binding affinity between the PDZ domain and ligand.

The flow cytometric screening strategy consisted of a positive selection using the fluorescent peptide ligands. Parameters were optimized to discriminate between

negative and positive controls. Outer cell membrane permeabilization, [85] as opposed to spheroplasting [86], was effective and the most efficient by eliminating the need to recover DNA plasmid after sorting. The optimized permeabilization buffer and reaction conditions that resulted in best cell survival and robust FACS signals were 50 mM Tris-HCl and 150 mM KCl with the pH adjusted to 7.4 buffer [94, 191] and a 30 minute reaction incubation at 4° C (see Figure 3 for a flow diagram of screening strategy).

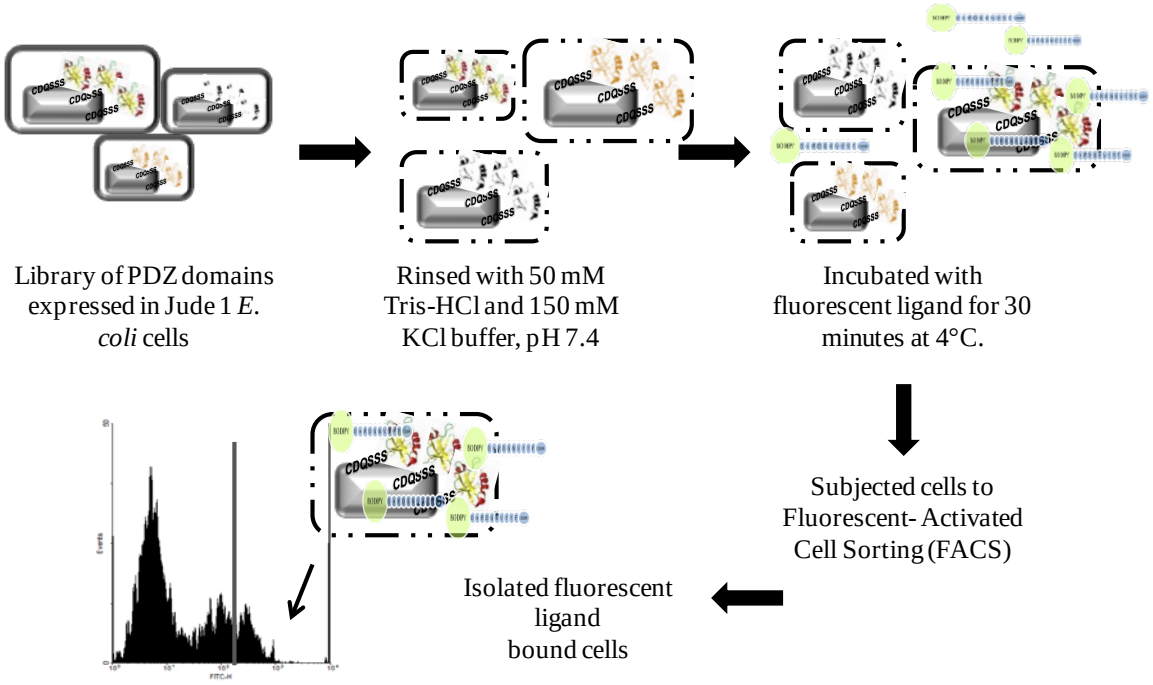


Figure 3. A library of proteins is expressed in the periplasm of *E. coli* cells and anchored to the periplasmic membrane. The cells are isolated by centrifugation to remove the growth media and resuspended and washed once with permeabilization buffer. The cells are incubated with fluorescent labeled peptide ligand, which can diffuse across the permeabilized, outer membrane. The cells are rinsed again with permeabilization buffer and then evaluated by FACS, isolating the fluorescing cells that bind the peptide ligand.

To verify the screening strategy was capable of discriminating between active and inactive PDZ binding domains consistently, controls were evaluated before each experiment. The inactive variant was constructed by replacing the four conserved

residues in the peptide ligand's *C*-terminus binding pocket with alanines. Flow cytometry confirmed this variant was inactive reflecting a fluorescent signal equal to cells with no PDZ plasmid expressed. The typical FACS profile of the controls depicted as the number of events (an event is a fluorescencing cell) versus the FITC-H (fluorescence intensity as interrogated by a laser with a wavelength of 500 nm) is shown in Figure 3.

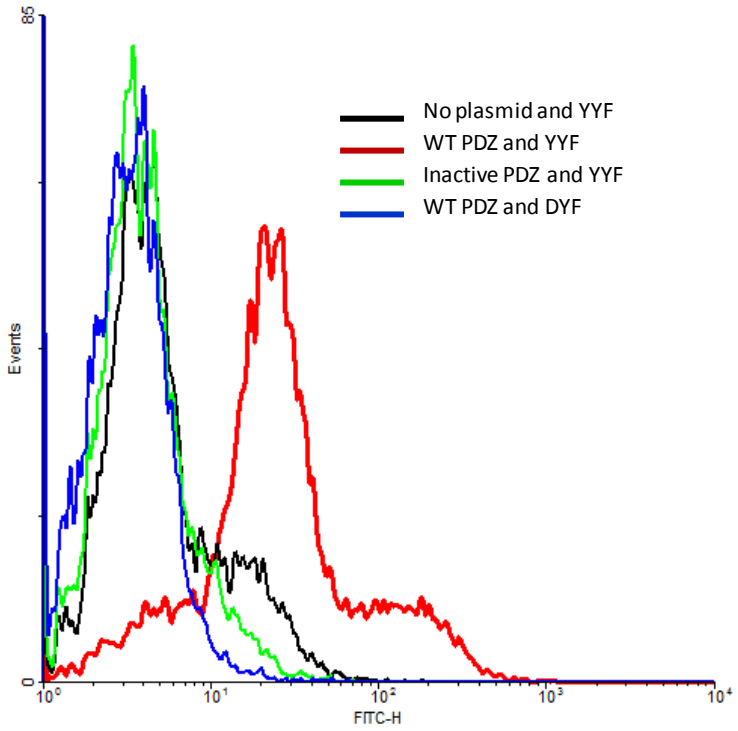


Figure 4. The typical fluorescence profile by flow cytometry discriminating between experimental controls. Fluorescence (FITC-H) signal on the x-axis versus events (number of fluorescing cells) on the y-axis for the control samples. Wild-type (WT) PDZ is the wild-type DegS PDZ domain. The inactive PDZ is the wild-type PDZ domain with four alanines replacing the four conserved residues in the peptide ligand's *C*-terminus binding pocket. YYF is the wild-type peptide ligand. DYF is the novel peptide ligand.

Library Construction

One random mutagenic library (named EP PDZ1) was constructed using error prone PCR conditions to amplify the parent DegS PDZ domain. After sequencing 10 clones, the EP PDZ1 library nucleotide mutation frequency was calculated to be 1.1% per PDZ domain. The EP PDZ1 contained 2.5×10^7 clones.

Eight libraries were constructed using oligionucleotide PCR primers that targeted two to five specific residues per library for mutagenesis. These targeted residues are hypothetically influential for peptide ligand binding based on evaluating the structure of DegS bound to wild-type peptide ligand and observing the proximal (within 4 angstroms) PDZ domain residues (see Figure 5). Two of these eight libraries were transformed to obtain libraries with 2×10^5 clones. The other six libraries were constructed but not transformed successfully. These libraries can be revisited.

The two libraries with targeted residues for mutagenesis are referred to as Rational Design PDZ 1 (RDP1) and Rational Design PDZ 2 (RDP2) (see Figure 6). In these two libraries, codons corresponding to five residues were mutagenized with an NNS codon, which encodes all 20 amino acid residues and minimizes stop codon frequency. Both libraries include residues G262, V283, T318, and M319 (highlighted in Figure 5). RDP1 also included residue V322 and RDP2 also included residue G263. Each of the ten clones sequenced from RDP1 and RDP2 had at least 4 out of 5 targeted randomized residues different from the wild-type codon in selected locations. All other codons retained the wild-type sequence as expected.

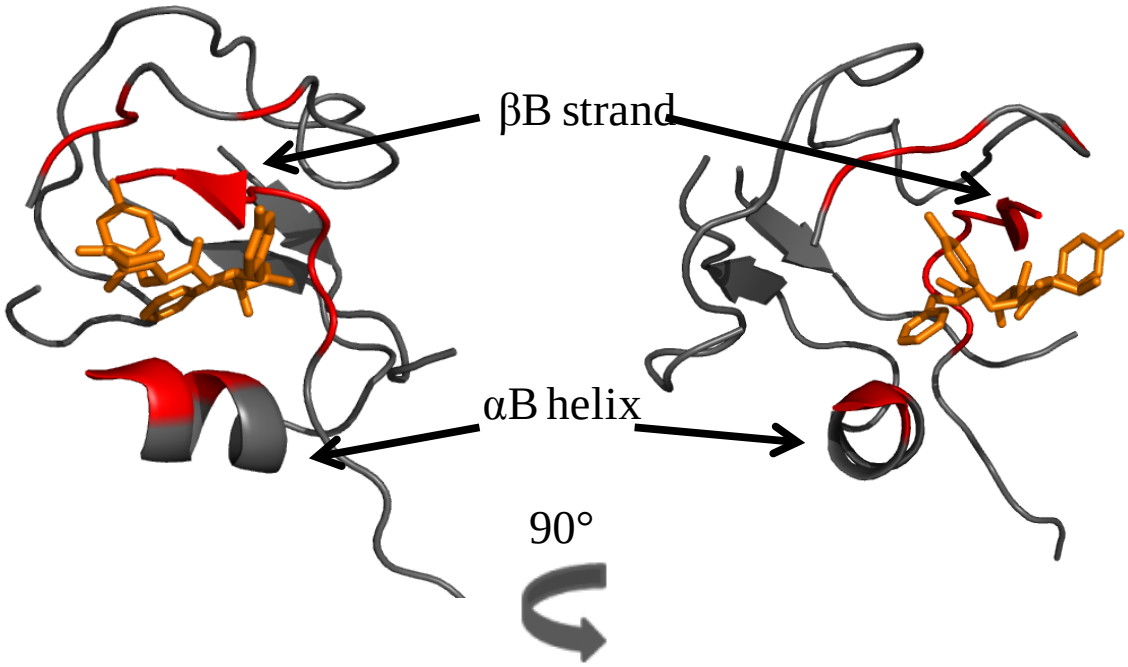


Figure 5. Two views of the wild-type DegS PDZ domain with wild-type peptide modeled in the binding pocket using PyMol. The ligand fits between a βB strand and αB helix characteristic of PDZ domain binding. This model reflects the most defined PDZ domain binding to ligand, yet still only has 67% of the PDZ residues defined. The peptide ligand is represented in stick configuration and is gold. The residues highlighted in red are within 4 angstroms of the peptide ligand (PDB: 1SOZ, [184]). Structures built with PyMol [104].

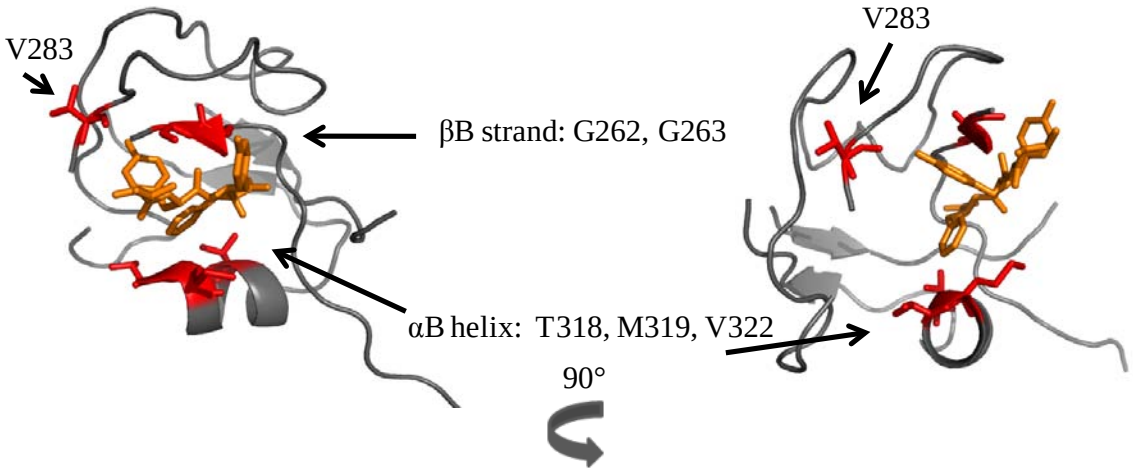


Figure 6. Two views of the wild-type DegS PDZ domain with the peptide ligand depicting residues mutated in libraries RDP1 and RDP2. The peptide ligand is represented by stick configuration and gold. The red residues represent the targets for mutation in the rationally designed libraries (RDP1 and RDP2). (PDB: 1SOZ, [184]). Structures built with PyMol [104].

Library Screening

The three libraries were each screened using the flow cytometric screening strategy previously described. Increasingly stringent screening conditions were employed for each successive round of screening, in that the fluorescent peptide ligand concentration was decreased from 1uM concentration to 0.25 uM for the final two rounds of FACS.

Library EP PDZ1 was screened with YYF and DYF separately. Ten clones isolated from the fourth round of screening EP PDZ1 with YYF were evaluated individually by FACS. Three clones had a significantly higher fluorescence mean than the wild-type PDZ domain (Figure 7, Panel A). Discouragingly, when screening the EP PDZ1 library with DYF, the first and second round library fluorescence profile overlapped the wild-type PDZ domain control fluorescence profile with DYF (Figure 7, Panel C). However, in the third round the fluorescence mean increased twenty-four-fold over the first two rounds. Ten clones from the fourth round, which had the highest fluorescence, were screened individually and all were found to have mean fluorescence intensity in excess of wild-type. These clones were also evaluated individually by FACS with YYF to determine if the DYF selected clones retained binding affinity towards the wild-type peptide ligand. Disappointedly, the clones retained and increased binding affinity towards wild-type peptide ligand, YYF. The three best clones selected after screening with DYF that were also the most selective as measured by FACS for DYF over YYF are shown in Figure 7, panel B. These six clones from EP PDZ1 were sequenced and evaluated further.

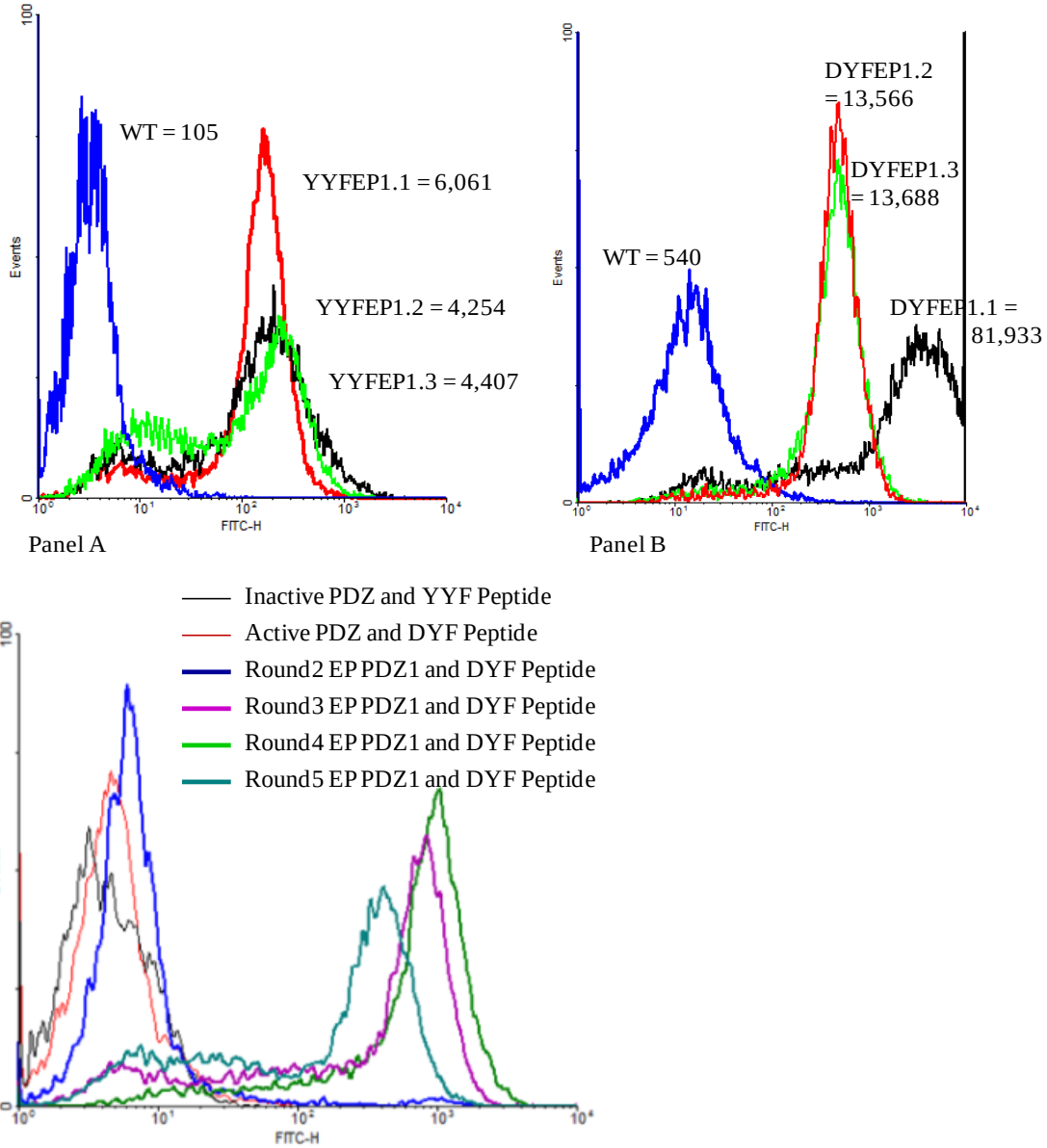


Figure 7. FACS data for sorting the random mutagenic library, EP PDZ1. Panel A depicts the fluorescence intensity profile (FITC-H, log x-axis) of the clones selected with wild-type ligand, YYF. The y-axis is the number of fluorescence events. Panel B the fluorescence intensity profile (FITC-H, log x-axis) of the clones selected with novel ligand, DYF. The y-axis is the number of fluorescence events. The blue profile labeled "WT" is the wild-type PDZ domain control with YYF peptide ligand in Panel A and B. The labeled values in Panel A and B are the FITC-H mean value for the individual clones. Panel C is the fluorescence intensity profile (FITC-H, log x-axis) of rounds 2 through 5 of EP PDZ1 sorted with the DYF peptide ligand.

The RDP1 and RDP2 libraries were screened as described for EP PDZ1 with peptides YYF and DYF. Interestingly, the RDP1 and RDP2 libraries converged to one sequence after five rounds for each peptide ligand screen, resulting in four unique clones. Surprisingly, the mutations did not correspond to the originally targeted mutations that were confirmed by sequencing before initiating sorting (see mutations listed in Figure 8 for each selected clone) indicating potential library contamination followed by the amplification of these clones through sorting. The sorting of the RDP1 and RDP2 libraries occurred simultaneously and after the sorting of the EP PDZ1 library.

Over all the screened libraries, two consensus mutations appeared simultaneously, N285I and M319K, in 4 out of 6 clones that were selected with peptide ligand DYF. The M319 residue was mutated to a basic residue in all six DYF selected clones. Mutants were constructed with an N285I or M319K mutation to probe the consensus mutations' contributions towards binding affinity. All of the selected clones, along with the constructed N285I and M319K clones, were inserted into the pGex6p vector for expression and further binding affinity characterization.

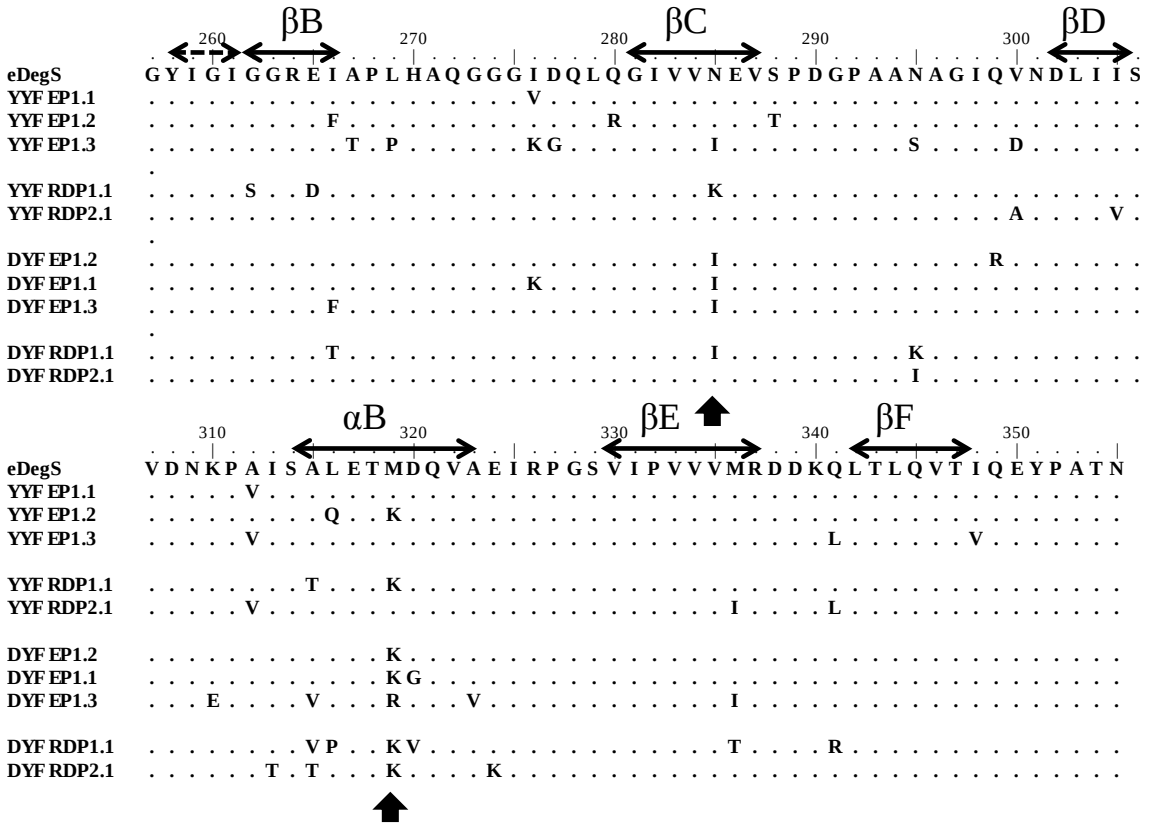


Figure 8. Summary of mutations in clones selected by FACS. All the clones are named with the last three amino acids of the peptide ligand used to select the clones (YYF is wild-type peptide ligand 1 and DYF is the novel peptide ligand 2), followed by the truncated library name (e.g. EP1 is EPDZ1 and RDP1 is RD PDZ1), and followed by a period and the clone number in sequential order of identification. The secondary structure features are indicated above the sequence. The dashed line with the double arrow indicates the c-terminal binding pocket. The solid, thick black arrows indicate consensus mutations. Figure generated by BioEdit.

PDZ domain Binding Analysis with Peptide ligands

The binding equilibrium coefficient (K_D) for each purified variant and wild-type DegS PDZ domain was quantified by fluorescence polarization measurements (Table 2). Binding analysis for the variants was performed in triplicate and the wild-type binding value represents an average of six measurements. The three variants selected using YYF from the EP PDZ1 library all had increased binding affinity (lower K_D values) towards

YYF. Interestingly, they had increased binding towards DYF as well. The three variants selected from the EP PDZ1 library using DYF had increased binding to DYF, but also had improved binding affinity for YYF. Variants isolated from the RDP1 and RDP2 libraries exhibited a similar trend. Only one variant, YYF RDP2.1, showed increased affinity towards YYF, which was used in its selection, without an increased binding affinity towards DYF. Interestingly, YYF RDP2.1 mutations were not concentrated in the αB or βB secondary structures like other variants (review mutations in Figure 8).

Table 2. Binding Affinity quantified by fluorescence polarization. The YYF column data are the equilibrium binding constants (K_D) calculated for the variant PDZ domain and YYF fit to the following equation: $mP_{Lmax}[PDZ]/K_D+[PDZ] = mP_L$ using nonlinear regression analysis for eight different PDZ variant concentrations. The DYF column data was calculated the same but using variant PDZ incubated with DYF. Error values are the standard deviation of the three data sets. ND indicates the binding affinity was not detectable. * indicates the data is not significant due to variance between measurements.

	Binding affinity (K_D) (μM)	
	YYF	DYF
WT	1.3 ± 0.2	ND
YYF EP1.1	0.15 ± 0.06	19.64 ± 3.52
YYF EP1.2	0.08 ± 0.03	0.83 ± 0.20
YYF EP1.3	0.42 ± 0.12	4.82 ± 1.53
DYF EP1.1	0.23 ± 0.08	0.85 ± 0.24
DYF EP1.2	0.06 ± 0.02	0.77 ± 0.17
DYF EP1.3	0.12 ± 0.04	0.94 ± 0.19
YYF RDP1.1	0.49 ± 0.08	9.02 ± 0.75
YYF RDP2.1	0.39 ± 0.02	ND
YYF RDP2.1.1	0.36 ± 0.13	2.08 ± 0.41
YYF RDP2.1.2	0.12 ± 0.03	4.57 ± 0.40
YYF RDP2.1.3	0.10 ± 0.03	3.55 ± 0.21
DYF RDP1.1	0.90 ± 0.46*	0.74 ± 0.52*
DYF RDP2.1	0.34 ± 0.08	5.16 ± 0.43
M319K	0.33 ± 0.30*	3.17 ± 0.31
N285I	0.29 ± 0.08	27.95 ± 5.10

Because the YYF RDP2.1 variant was selective for increased binding to YYF, a second generation random mutagenesis library was constructed using YYF RDP2.1 as the parent template with error prone PCR conditions. A library of 5×10^7 clones with a nucleotide error rate of 2.4%, based on the sequencing of 10 randomly chosen clones, was made. After five rounds of FACS sorting using peptide ligand YYF, ten individual clones were analyzed individually. The best three variants were sequenced

(named YYF RDP2.1.1 through YYF RDP2.1.3 to indicate their parent template YYF RDP2.1) and inserted into the pGEX6p plasmid for expression and purification, and their binding affinity quantified by fluorescence polarization. The three variants had a consensus N285I mutation (see Table 3). Binding affinity for the YYF ligand increased by three-fold over the parent YYF RDP2.1 variant for two of the second generation variants, however, this was accompanied by the acquisition of affinity for DYF, thereby abrogating the specificity advantage of the parent YYF RDP2.1 clone (Table 2).

Table 3. Mutation summary of second generation random mutagenesis library using YYF RDP2.1 as the parent template for amplification in error prone PCR conditions.

Variant Name	No. of Mutations																										
Secondary Structure →		βB							βC		αA		βD		αB								βE				
Amino Acid Position →		262	263	265	266	267	272	276	285	288	290	300	305	312	313	314	315	316	319	320	323	329	336	339	340	341
WT		Gly	Gly	Glu	Ile	Ala	Gln	Ile	**Asn**	Ser	Asp	Val	Ile	Ala	Ile	Ser	Ala	Leu	Met	Asp	Ala	Ser	Met	Asp	Lys	Gln
YYF RDP2.1	5											Ala	Val	Val									Ile			Leu
YYF RDP2.1.1	5						Leu		**Ile**	Pro	Asn											Thr				
YYF RDP2.1.2	5							Leu	**Ile**							Tyr								Gly		Leu
YYF RDP2.1.3	2								**Ile**																Arg	

The affinity change for each variant is the ratio of wild-type DegS PDZ binding affinity to variant PDZ domain binding affinity rounded to the nearest five (or integer if less than 5) to provide a qualitative perspective of relative affinity change amongst the PDZ variants. The statistical significance of the difference between each variant binding affinity and the wild-type binding affinity was determined using a t-test and converting the t-value to a p-value. The binding difference is significant for p values less than 0.05, which means there is a 95% confidence level that the two values are significantly different. DYFEP1.1 had the highest relative affinity change with a 20-fold decrease in K_D for YYF over wild-type K_D. Interestingly, the group of variants selected from the EP

PDZ1 library with the DYF peptide ligand has the highest affinity increase towards YYF peptide ligand (see Table 4).

Table 4. The PDZ variants affinity change towards the wild-type YYF peptide ligand relative to the wild-type DegS PDZ K_D. [1] Affinity change is the ratio of WT:Variant K_D for YYF ligand. [2] The p value is the probability the WT and the PDZ domain affinities are different. The affinities are significantly different for p values < 0.05. NS indicates the data was not significantly different from the WT K_D.

	Binding affinity (K_D) (μM)		
	YYF	Affinity Change[1]	p value[2]
WT	1.3 ± 0.2		
YYF EP1.1	0.15 ± 0.06	10	<0.0001
YYF EP1.2	0.08 ± 0.03	15	<0.0001
YYF EP1.3	0.42 ± 0.12	3	0.0002
DYF EP1.1	0.23 ± 0.08	5	<0.0001
DYF EP1.2	0.06 ± 0.02	20	<0.0001
DYF EP1.3	0.12 ± 0.04	10	<0.0001
YYF RDP1.1	0.49 ± 0.08	3	<0.0001
YYF RDP2.1	0.39 ± 0.02	3	<0.0001
YYF RDP2.1.1	0.36 ± 0.13	4	0.0002
YYF RDP2.1.2	0.12 ± 0.03	10	<0.0001
YYF RDP2.1.3	0.10 ± 0.03	10	<0.0001
DYF RDP1.1	0.90 ± 0.46*	NS	0.318
DYF RDP2.1	0.34 ± 0.08	4	<0.0001
M319K	0.33 ± 0.30*	4	0.014
N285I	0.29 ± 0.08	4	<0.0001

The specificity of each PDZ variant is the ratio of variant PDZ and YYF binding affinity to variant PDZ domain and DYF binding affinity rounded to the nearest five to provide a qualitative perspective of relative specificity amongst the PDZ variants. None of the PDZ variants were more specific for the novel peptide ligand DYF than wild-type ligand YYF; thus, specificity is in terms of the wild-type peptide ligand, YYF. The

higher specificity values indicate the more specific the PDZ variant is for YYF. The wild-type PDZ domain and YYF RDP2.1 are most specific towards YYF since there was no detectable binding for DYF. Interestingly, the M319K variant binding affinity towards DYF was nine-fold better than the N285I variant (see Table 5). Generally, as the variant's binding affinity was improved for DYF, it was also improved for YYF.

Table 5. The PDZ variants specificity towards the wild-type YYF peptide ligand relative to the novel peptide ligand, DYF. [1] Specificity is the ratio of DYF K_D:YYF K_D. [2] The p value is the probability the DYF K_D and the YYF K_D are different. The affinities are significantly different for p values < 0.05. NS indicates the data was not significantly different from the WT K_D. ND indicates binding was not detectable ‡ indicates the specificity is specific for YYF, since the DYF binding was not detectable.

	Binding affinity (K_D) (μM)			
	YYF	DYF	Specificity[1]	p value[2]
WT	1.3 ± 0.2	ND	‡	
YYF EP1.1	0.15 ± 0.06	19.64 ± 3.52	130	0.001
YYF EP1.2	0.08 ± 0.03	0.83 ± 0.20	10	0.024
YYF EP1.3	0.42 ± 0.12	4.82 ± 1.53	10	0.008
DYF EP1.1	0.23 ± 0.08	0.85 ± 0.24	NS	0.061
DYF EP1.2	0.06 ± 0.02	0.77 ± 0.17	10	0.019
DYF EP1.3	0.12 ± 0.04	0.94 ± 0.19	5	0.020
YYF RDP1.1	0.49 ± 0.08	9.02 ± 0.75	20	0.003
YYF RDP2.1	0.39 ± 0.02	ND	‡	
YYF RDP2.1.1	0.36 ± 0.13	2.08 ± 0.41	5	0.024
YYF RDP2.1.2	0.12 ± 0.03	4.57 ± 0.40	10	0.003
YYF RDP2.1.3	0.10 ± 0.03	3.55 ± 0.21	35	0.001
DYF RDP1.1	0.90 ± 0.46*	0.74 ± 0.52*	NS	0.803
DYF RDP2.1	0.34 ± 0.08	5.16 ± 0.43	15	0.003
M319K	0.33 ± 0.30*	3.17 ± 0.31	10	0.015
N285I	0.29 ± 0.08	27.95 ± 5.10	95	0.011

DISCUSSION

Building upon the collective lab experience derived from engineering antibody binding affinity, we were able to establish a high throughput flow cytometric screening method to engineer increased binding affinity into the DegS PDZ domain. The method depended on expression of an accessible, active PDZ domain coupled to retention of the target peptide ligand in analogy to several previously reported antibody engineering approaches [85, 86]. A fluorescence anisotropy based assay was used to quantify selected DegS variant binding affinities.

Increased affinity for wild-type peptide

Using only a positive selection criterion, several DegS variants were isolated with improved affinity for the wild-type YYF peptide ligand after four to five rounds of sorting. Despite being selected using the novel peptide ligand DYF, the highest affinity variant was DYF EP1.2, which has an equilibrium dissociation constant (K_D) of 60 nM (see Table 4) with the wild-type peptide ligand YYF, reflecting a 20-fold affinity enhancement.

PDZ Domain Class Switch

Our second objective was to change the affinity of the PDZ domain from a Class II (YYF, the wild-type ligand) to a Class III peptide ligand (DYF). In some sense, this objective was achieved. Interestingly, the best variant for the Class III peptide ligand was the same variant, DYF EP1.2 that exhibited the highest affinity for the wild-type Class II peptide ligand. DYF EP1.2 has an equilibrium dissociation constant (K_D) 770 nM for the Class III peptide ligand, DYF. This result is consistent with other PDZ engineering studies and confirms that PDZ domains can also be customized to bind different ligands using *E.coli* cell-based selection strategies [63, 64, 70, 73, 137].

Overall Lack of Specificity

All of the variants with the highest binding affinities displayed relaxed specificity. This disappointing result highlights the disadvantage of using positive selection criteria alone in the screening strategy. In particular, all of the high affinity variants resulting from the DYF ligand screening had equivalent or better affinity to the wild-type YYF ligand (see Table 5).

Consensus Mutation Analysis

The relative affinities of the two constructs containing a single consensus mutation were intuitively consistent with the library screening results. The M319K variant construct enhanced binding to the DYF peptide ligand more than the N285I variant construct with a K_D equal to 3.17 ±0.31 μM compared to the N285I K_D of 27.95± 5.10 μM. M319 was mutated to a basic residue in all the DYF selected variants. Variants with the N285I consensus mutation were found in selected variants from both peptide ligand screens, suggesting the N285I mutation impacted binding affinity through a ligand position other than the (-2) ligand residue.

M319 is located in the PDZ domain αB:5 position and N285I is located in the βC:4 position. Tonikian *et al.* evaluated 82 human and *C. elegans* PDZ domains to determine the relationship between the PDZ sequence and ligand binding specificity [136]. The consensus mutation binding affinity results from this study corroborate their data that suggests the PDZ residue in the αB:5 position impacts PDZ binding to the (-2) residue in the peptide ligand primarily. Additionally, their results indicate the PDZ residue in βC:4 position impacts binding to the (-1) and (-3) peptide ligand positions. Our data suggest that N285I, the residue in the βC:4 position, impacts binding to peptide ligand in a position other than the (-2) peptide position; thus, it is also consistent with Tonikian *et al.*'s results.

Structural Analysis

The mutations were examined in the context of the published DegS PDZ domain structures. All the DegS structures in the protein database include the protease domain, which was not included in this analysis for clarity. The best DegS PDZ domain structure bound to peptide ligand has 67% of the residues resolved (PDB: 1SOZ [184]) and is used in this analysis to evaluate potential effects of specific mutations. The unbound DegS PDZ domain is more complete (97%) and is a better representation of the PDZ secondary structure elements. As a result, the unbound DegS PDZ domain is used to model all the mutations found in the selected variants in Figure 9 with the consensus mutations, M319K and N285I, highlighted in red (PDB: 1TEO[184]) using PyMol to select the most probably rotamer without steric hindrances. As expected, mutations are concentrated in the αB helix and βB strand that form the peptide ligand binding groove. Additionally, mutations were concentrated in the βC strand and the loop between the βB and βC strands. Other mutations were found scattered throughout the PDZ domain and it is inconclusive whether these scattered mutations were simply non-deleterious for PDZ stability and ligand binding or positively impacted ligand binding.

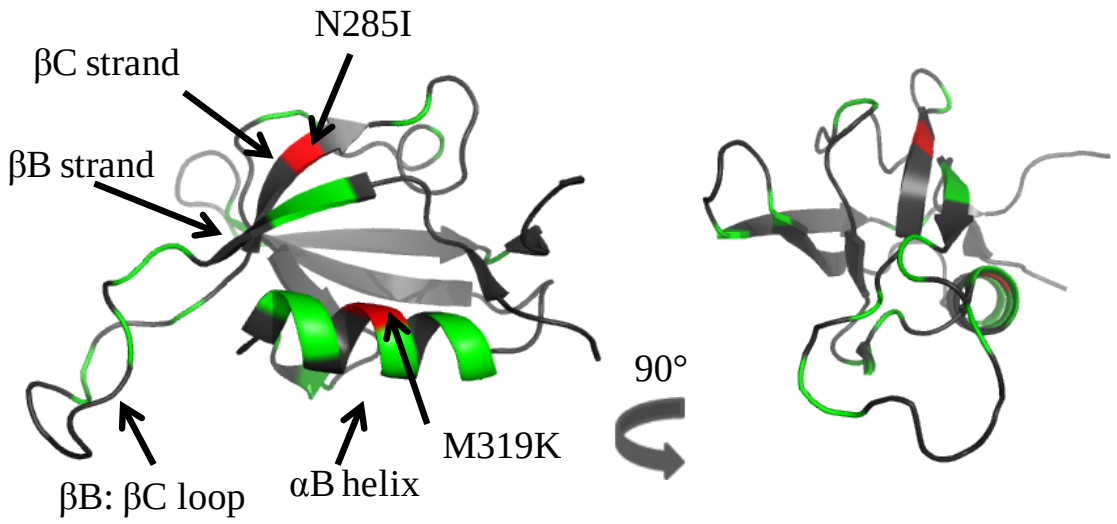


Figure 9. Two views of the DegS PDZ domain unbound structure modeling all the mutations (in green) represented by the selected variants from both screens. The secondary structural elements are pointed out that contain the highest concentrations of mutations. The two consensus mutations are also labeled and shown in red. (PDB: 1TEO, [184]). Structures built with PyMol [104].

The variants isolated from screening with the Class III ligand, DYF, had two simultaneous consensus mutations: N285I and M319K. Position 319, located at the αB:5 postion, is a peptide ligand contact residue within the PDZ binding groove, specifically located near peptide ligand position (-2). The M319K mutation appears to stabilize the binding of the Class III ligand through a potential interaction between the basic residue, Lys, and the acidic peptide ligand residue, Asp. The PyMol model in Figure 10, shows the most probable rotamer of the mutant residue, Lys, is within 2 angstroms of the (-2) peptide residue, Asp, permitting a polar interaction. The effect of the N285I mutation in position βC:4 is not obvious in the structure as it does not make a direct contact with the peptide ligand as it appears the βB strand is between the peptide ligand and the βC strand. However, it is difficult to discern due to the lack of structure resolution of residues in and near the βB strand.

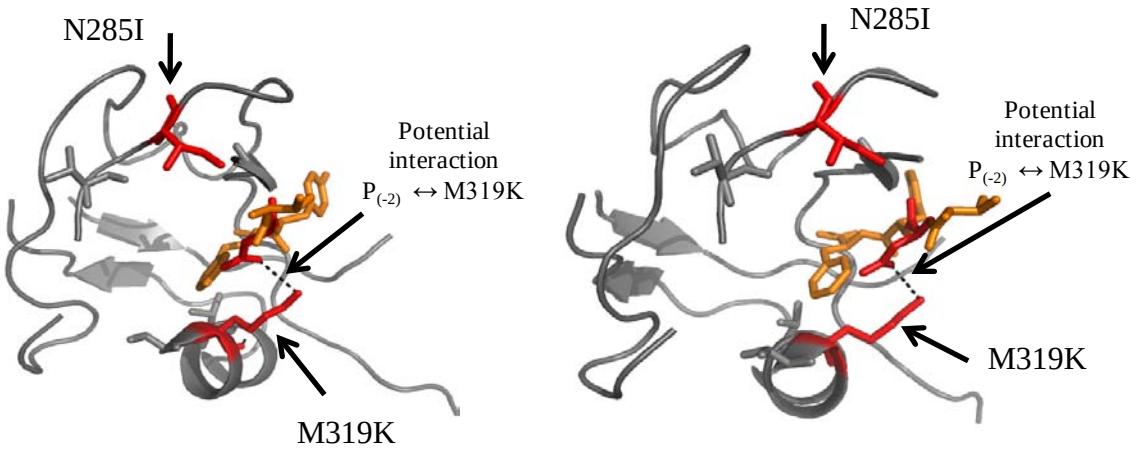


Figure 10. Two views of the Class III ligand bound to the DegS PDZ domain are modeled (PDB: 1SOZ, [184]) with consensus mutations inserted as the most probable rotamer. The $P_{(-2)}$ ligand residue is highlighted red and in stick configuration, the rest of the peptide ligand is in gold. The PDZ residues highlighted in red are consensus mutations, N285I and M319K, from the Class III ligand screen. The potential polar contact between M319K with the ligand is represented with a black dashed line. Structures built with PyMol [104].

Because all the isolated variants resulted in improved affinity towards the wild-type preferred Class II ligand YYF, the mutations that proved consensus for the Class III ligand variants must also create an enhanced binding groove for the YFF peptide. Interestingly, the variants selected when using the wild-type ligand YYF did not include variants with simultaneous N285I and M319K mutations like the variants selected using the novel peptide ligand, DYF.

CONCLUSION

An *E. coli* cell-based screening methodology was successfully developed to improve the binding affinity of a PDZ domain. The flow cytometry screening method is straight forward and efficient. Five rounds of screening can be completed in nine days. Successful variants were found in three to five rounds with up to a 20-fold increased

binding affinity to the wild-type preferred ligand. Significantly, the binding affinity to a novel peptide ligand was also created in five rounds of sorting. The engineered PDZ selectivity, however, was relaxed, accommodating both peptide ligands.

The screening methodology may be improved by incorporating a negative selection criterion and improving the targeted mutagenic library design and using improved method technique. A negative selection criterion can be incorporated by using two selection peptide ligands, labeled with different and distinguishable fluorophores. The cells can then be screened by rejecting a group of cells binding to the negative selection peptide and collecting the cells binding preferentially to the positive selection peptide. Two fluorophores have been used successfully in this lab for enzyme engineering [192]. The targeted mutagenic library approach needs to be revisited with improved technique and targeting less PDZ residues to more methodically determine PDZ residue binding impact on specific ligand positions.

This methodology has potential applications as an alternative to engineering PDZ domains through phage display or the yeast two-hybrid screen as previously done. Redesigning PDZ domains to bind with high affinity and specificity may be useful for biotechnological applications that require small binding proteins viable in reducing and non-reducing environments. Many viruses including HPV (a cause of cervical cancer) and H5N1 (avian flu) encode proteins that bind with a cellular PDZ domain leading to virulence [193, 194]. These proteins' PDZ binding motifs have been targeted for designing assays to detect the virulent viral strains and also investigated as a potential therapeutic target (Arbor Vita Corporation, San Diego, CA).

CHAPTER 3: LINKING ENZYME ACTIVITY TO PEPTIDE DETECTION

CHAPTER SUMMARY

Objective: The objective is to engineer an allosteric protein by fusing its protein binding domain that was redesigned by laboratory evolution to its wild-type protease domain to create a novel or more sensitive "switch" that will activate protease activity.

Approach: The well-characterized allosteric *E. coli* protein, DegS, was used as the template for allosteric protein engineering. DegS peptide ligand binding modules, PDZ domains, previously engineered to bind to the wild-type activation peptide ligand with up to 20-fold greater affinity than the wild-type PDZ domain and engineered to bind to a novel activation peptide ligand, were fused to the wild-type DegS protease domain. These DegS variant enzymes were then evaluated for allostery with the wild-type and novel activating peptide ligands by two methods: a conventional SDS-PAGE assay and a chromogenic enzyme product assay using a substrate peptide conjugated to *p*-nitroanaline.

Results: Four engineered PDZ domains were fused to the DegS protease domain and evaluated for allosteric activation with the two peptide ligands. None of the DegS enzyme variants were equal to wild-type DegS protease activity with the wild-type activating peptide ligand. Still, one of the DegS enzyme variants, YYF EP1.1, did retain a degree of allostery and was activated by the wild-type activating peptide ligand for a 2-fold protease activity increase over baseline activity. YYF EP1.1 also gained the same degree of allostery with the novel activating peptide ligand.

INTRODUCTION

Proteins are engineered by laboratory evolution techniques that include bacterial and yeast cell display for more than one characteristic at a time for instance expression and binding affinity or folding and binding affinity [88, 89, 91]. The two characteristics are related such that a protein cannot bind its ligand unless folded properly or two fluorophores (one fluorophore linked to an expression marker and the other linked to the binding ligand) are used to set multiple sort parameters for fluorescent activated cell sorting (FACS). These techniques were used to increase antibody binding affinity for

potential therapeutic efficacy against anthrax toxin while improving expression or folding without disulfide bonds to facilitate antibody production [88, 89]. On the other hand, there are no examples of simultaneously engineering the two linked activities of an allosteric protein. Instead, redesigning an allosteric protein is undertaken by performing iterations of combining protein modules from different proteins or circular permutations of protein modules until a combination is found that links the two activities desired such as a binding module and a proteolysis module to create a proteolytic "switch" that is activated when the activating ligand binds the binding module [156, 195, 196].

In this study, we investigate if previously engineered peptide binding modules of an allosteric protein fused with the wild-type protease domain will form an allosteric protein with a novel or more sensitive activating "switch". The allosteric protein used in this study is the well characterized *E. coli* periplasmic membrane anchored serine endoprotease, DegS, which is activated by the *C*-termini of unfolded outer membrane porins (OMPs), which bind to the DegS PDZ domain. The DegS PDZ domain can recognize a variety of *C*-terminal sequences as long as they conform to the conventional Class II peptide ligand sequence consensus (Φ-X-Φ, with X denoting any L-amino acid and Φ representing hydrophobic amino acid residues) and consist of at least a tri-peptide. The molecular mechanism of DegS allostery has been investigated and ten crystal structures of the full length protein (including variants) have been solved, two structures with unbound ligand and eight structures with bound activating ligand to the DegS PDZ domain [181-185, 187].

Studies suggest that increased ligand binding affinity to the PDZ domain results in increased activation of the protease domain; however, the molecular dynamics of how the DegS protein adjusts for different activating peptide ligands is not fully understood. Clausen *et al*. suggest the "Peptide Activation Model", which hypothesizes the degree of

DegS proteolysis activation is controlled by the activating peptide ligand residue in position (-1) interacting directly with Loop 3 in the protease domain [181]. Sauer *et al.* suggest the "Inhibition Relief Model", which hypothesizes the peptide ligand residue in position (-1) does not determine proteolysis rates but activating peptide ligands with higher PDZ domain binding affinity activate proteolysis faster by relieving the PDZ domain inhibition of the proteolytic domain, which then shifts the equilibrium of DegS towards the active form that preferentially binds the RseA substrate. Both models posit that DegS can recognize different misfolded proteins in the periplasm with Class II characteristic *C*-termini and can then respond with an appropriate proteolytic rate based on the particular misfolded protein concentration [168, 186].

Both groups' biochemical and structural data support their respective DegS activation mechanism hypothesis and yet does not disprove the other hypothesis. Both groups have crystallized DegS with different mutations bound to different peptide ligands; however, the structure data are inconclusive. The structures of ligand bound PDZ domains have low resolution with 40 to 80% of the residues defined. Protease domain Loop 3 conformation even varies amongst the subunits of the DegS trimers crystallized making it difficult to conclude if the peptide ligand residue in position (-1) is influencing the Loop 3 conformation [34, 181, 183, 184, 187]. Understanding the molecular mechanism of DegS activation would assist in redesigning DegS. At this point, redesigning DegS may contribute to a better understanding of the molecular mechanism.

In this investigation, six variant DegS PDZ domains were fused to the wild-type DegS protease domain and these fusions were evaluated for allosteric activation of the protease domain with two different 10-mer activating peptide ligands (DNRDGNVYYF and DNRDGNVDYF). Four of the variant PDZ domains were selected from a random

mutagenic combinatorial library of 2.5 x 10^7 clones that was expressed in the periplasm of *E. coli* cells by flow cytometry with FACS. Two of the variants (YYF EP1.1 and YYF EP1.2) were selected for improved binding to the wild-type Class II peptide ligand, DNRDGNVYYF (referred to as YYF), and two variants (DYF EP1.1 and DYF EP1.2) were selected for binding to Class III peptide ligand, DNRDDGNVDYF (referred to as DYF). Binding between the Class III peptide ligand and the wild-type DegS PDZ domain was not detectable. The two additional PDZ domains used in this study were constructed with single consensus mutations found amongst the four selected clones and these PDZ domain constructs are referred to as: N285I and M319K.

MATERIALS AND METHODS

Cloning DNA constructs

The DegS fusion protein sequences were constructed by overlap PCR using the DegS protease domain and the variant PDZ domains that were initially prepared separately. The DegS protease domain (residues 27 through 257) minus the transmembrane and PDZ domains and incorporating the Nhe1 restriction enzyme site at the 5' end was amplified from genomic *E. coli* DNA using a standard PCR protocol with the forward DNA oligo: 5' aaaaaagctagccgcagccttaacccgctttcc; and the reverse DNA oligo: 5' cgtccgccgataccaatgtagccgcggatcacgcgaccatc. All the PDZ variant domains, in this investigation, were amplified from the clone's pAPEX (used in FACS screening) or pGEX6p plasmid (used in PDZ protein purification for fluorescence polarization measurements), using a standard PCR protocol with the forward DNA oligo: 5' ggctacattggtatcggcggacg; and the reverse DNA oligo that incorporated the XhoI restriction enzyme site at the 3' end: 5' cgtccgccgataccaatgtagccgcggatcacgcgaccatc. The

DegS protease domain and PDZ domain construct sizes were confirmed by gel electrophoresis, which was visualized with ethidium bromide. These constructs were purified using a standard DNA purification protocol (Qiagen, Germany) and then added together to a standard PCR reaction where the overlapping sequence between the DegS protease domain and the PDZ domain annealed to make the parent DNA template for the fusion protein. The PCR reaction contained the same forward oligo used in the DegS protease domain PCR amplification reaction and the same reverse oligo used in the PDZ variant PCR amplification reaction. The DegS fusion product sizes were confirmed by gel electrophoresis, purified, and then incubated with the restriction enzymes, NheI and XhoI (NEB, Ipswich, MA). The pet28a plasmid was incubated separately with the same restriction enzymes. The restriction enzyme products were gel purified by standard DNA preparative methods (Zymo Research, Orange, CA). The linearized pet28a plasmid and the DegS fusion clones were then ligated together using Quick T4 ligase (NEB, Ipswich, MA) and chemically transformed into BL21(DE3) cells. The sequences of selected clones were confirmed by the ICMB DNA sequencing facility and it was also verified the *N*-terminal his tag from the pet28a plasmid was in frame with the cloned sequence.

Additionally, the DegS protease substrate, RseA periplasmic domain (residues 121-216), was amplified by a standard PCR protocol from genomic *E. coli* DNA that incorporated the NheI and XhoI restriction enzyme sites at the 5' and 3' ends respectively using the forward DNA oligo: 5'aaaaaagctagctataatggacaatctgaaacgtcccagcagc and the reverse DNA oligo: 5' aaaaaactcgagttattactgcgattgcgttcctaaagtttgaattcc. The RseA clone was prepared the same as above, incubating with NheI and XhoI restriction enzymes (NEB, Ipswich, MA) and gel purified to prepare for ligation into the pet28a plasmid linearized with the same restriction enzymes. The ligation product was chemically transformed into BL21 (DE3)

cells. The sequences of selected RseA clones were confirmed by the ICMB DNA sequencing facility and it was also verified the *N*-terminal his tag from the pet28a plasmid was in frame with the cloned sequence.

Protein purification

The wild-type periplasmic RseA domain, and the wild-type DegS and DegS fusions without their trans-membrane and signal domains were cloned into the pET28a plasmid as described above such that the protein sequence was preceded with a 6XHis tag at the *N*-terminus. The sequence confirmed clones were transformed into BL21(DE3) cells. All proteins were expressed and purified using the same method.

Cultures were grown overnight to saturation at 30 ºC in 2xYT media supplemented with 100 ug/ml ampicillin and 2% glucose from a single colony. The saturated cultures were then sub-cultured into a 1:100 dilution of supplemented 2xYT media. The culture was grown to an OD_{600} of 0.5 to 0.7. At this point, the culture was induced for expression with 0.1 mM of IPTG for 3 hours to overnight. The cell pellet was spun down and the media, the supernatant, decanted. The cell pellet was stored at -20 ºC for one to four days. The cell pellet was resuspended in 10% of the original culture volume with Lysis Buffer (PBS (50mM Na_2HPO_4 (pH =8.0), 300mM NaCl) with 10mM imidazole chilled at 4 ºC) and the cell membraness were disrupted using a French pressure cell. The lysate was then centrifuged at 16,000 rpm (30,000*xg*) in the Beckman JA20 rotor to separate the soluble and insoluble fractions. The soluble fraction was applied to 1% of the original culture volume of Ni-NTA beads (Qiagen, Germany) packed in a 5 ml disposable column and equilibrated with Lysis buffer. The soluble fraction was passed through the column once by gravity flow. The column was then washed with 10% of the original culture volume of Wash Buffer (PBS with 20mM imidazole chilled at 4 ºC) twice by gravity flow. After washing, 1.5X resin volume of

Elution Buffer (PBS with 250mM imidazole chilled at 4 ºC) was used to elute the protein. The eluted protein was then dialyzed 1:1000 volume ratio into storage buffer (150 mM NaH_2PO4, pH 8.1, 380 mM NaCl, 10% glycerol, 0.4mM EDTA). The collected product was then concentrated to 200 to 500 µM. Purity was confirmed by 16% SDS-PAGE gel visualized with gelCodeBlue reagent (Thermo Scientific, Rockford, IL). Concentration was determined with a BCA assay (Thermo Scientific, Rockford, IL) using BSA as a standard and analyzing all samples in duplicate.

Cleavage Assay evaluated by gel electrophoresis

First 0.5 to 5µM of wild-type (WT) DegS or the appropriate variant were incubated with 30 to 60 µM of activating ligand for 10 minutes at 37 °C in the DegS reaction buffer established by Sohn *et al*.(150 mM NaH_2PO_4, pH 8.1, 380 mM NaCl, 10% glycerol, 0.4mM EDTA) [182]. 100 µM of purified RseA periplasmic domain was added. The reactions were incubated for 1-3 hours at 37 °C. Reactions were stopped by adding one sample volume of 2X SDS loading buffer and freezing. Samples were electrophoresed on 16% Tris Glycine gels (pre-cast from Invitrogen, Carlsbad, CA). Proteins and cleavage products were visualized after staining by gelCodeBlue Reagent (Thermo Scientific, Rockford, IL).

***p*-Nitroanaline Assay**

Protease assays using the synthetic substrate *N*-methoxysuccinyl-AAPV-*p*-nitroanaline (Sigma Aldrich, St Louis, MO) were performed at 37 °C in the same DegS reaction buffer (150 mM NaH_2PO_4, pH 8.1, 380 mM NaCl, 10% glycerol, 0.4mM EDTA) with 3 to 12µM of purified DegS or variant and 2-4 mM of substrate by measuring the changes in OD_{410} with the Synergy spectrophotometer (GE Corporation, Fairfield, CT) continuously for 1 h and for a single point after 12-16 hours (overnight).

Varying amounts of activating peptides were pre-incubated (37 °C) with DegS for 10 minutes before adding the ρ-NA substrate. The specific activity was calculated by the following equation.

Specific Activity (nmol $mg^{-1}min^{-1}$) = $OD_{410}*V_{final}$ /(□ *l*mg_D*min)

Abs_{410} = absorbance measured at 410 nm,

V_{final} = final reaction volume (L),

mg_D = mg of DegS,

□ = substrate extinction coefficient (8,800 L/mol*cm), and

min = reaction time in minutes,

l = length path of spectrophotometer (cm).

RESULTS

PDZ domains selected for DegS fusion proteins

The four selected variant PDZ domains fused to the wild-type DegS protease domain all had improved affinity for the wild-type peptide ligand, YYF, with equilibrium dissociation constants (K_D) measured by fluorescence polarization as reported in Chapter 2 of this dissertation. Additionally, the two PDZ domain constructs with single consensus mutations, which are used to discern the impact of mutations in the selected variants, also had improved binding affinity to both peptide ligands. DYF EP1.2, despite being selected using the Class III peptide ligand, had 20-fold higher affinity for wild-type peptide ligand, YYF, than the wild-type DegS PDZ domain binding affinity. YYF EP1.1 retained the most selectivity for the wild-type peptide ligand, YYF, over the DYF peptide ligand but still had improved affinity for the DYF peptide ligand over wild-type DegS (see Table 1).

Table 1. Binding Affinity quantified by fluorescence polarization. The YYF column data are the equilibrium binding constants (K_D) calculated for the variant PDZ domain and YYF fit to the following equation: $mP_{Lmax}[PDZ]/K_D+[PDZ] = mP_L$ using nonlinear regression analysis for eight different PDZ variant concentrations. The DYF column data was calculated the same but using variant PDZ incubated with DYF. Error values are the standard deviation of the three data sets. ND indicates the binding affinity was not detectable. * indicates the data is not significant due to variance between measurements.

	Binding affinity (K_D) (μM)	
	YYF	DYF
WT	1.3 ± 0.2	ND
YYF EP1.1	0.15 ± 0.06	19.64 ± 3.52
YYF EP1.2	0.08 ± 0.03	0.83 ± 0.20
DYF EP1.1	0.23 ± 0.08	0.85 ± 0.24
DYF EP1.2	0.06 ± 0.02	0.77 ± 0.17
M319K	0.33 ± 0.30*	3.17 ± 0.31
N285I	0.29 ± 0.08	27.95 ± 5.10

The mutations of the selected four clones are shown in Figure 1 with the consensus mutations, N285I and M319K, underscored. The mutations occur simultaneously in DYF EP1.1 and DYF EP 1.2. M319K occurs in three out of the four variants. YYF EP1.1 is the only variant that does not include a consensus mutation.

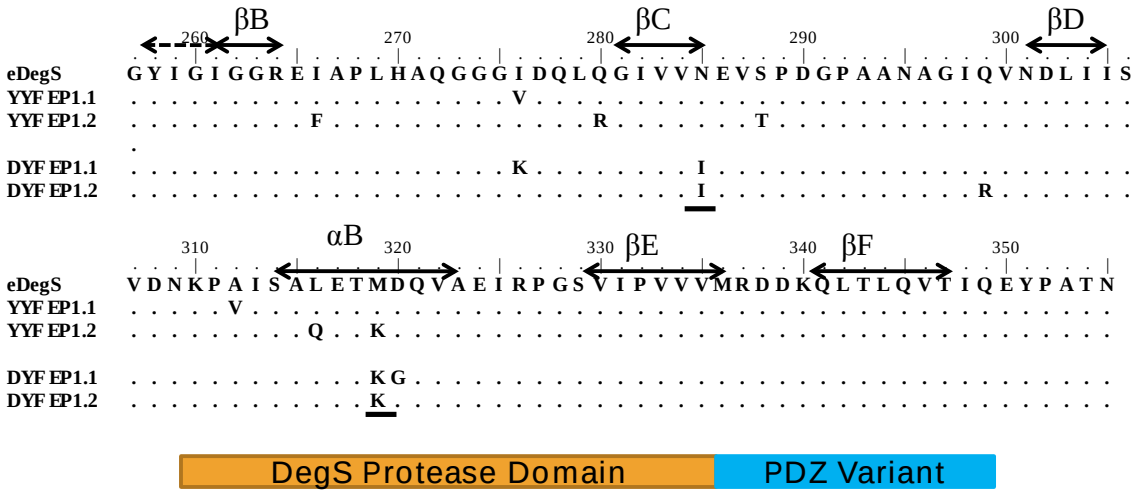


Figure 1. PDZ Variants' Alignment with the wild-type DegS PDZ domain. The consensus mutations, N285I and M319K, are underlined. The alignment was generated by BioEdit. YYF EP1.1 has two mutations, YYF EP1.2 has five mutations, DYF EP1.1 has four mutations and DYF EP1.2 has three mutations. These PDZ domains were fused with wild-type DegS protease domain as shown by the bar diagram to create a full length variant protease.

The objective was to determine if the protein fusions that are comprised of the DegS wild-type protease domains and the four selected variant domains are allosteric. A conventional sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) assay was used to determine if the protein fusions were qualitatively allosteric [168, 172, 180, 181]. Additionally, a second chromogenic product assay was applied to further probe allostery using a gradient of activating ligand concentrations.

Protein Gel Analysis

In vitro reactions consisted of 2 µM purified soluble DegS, 60 µM activator, and 100 µM RseA were prepared in DegS reaction buffer (150 mM NaH_2PO_4, pH 8.1, 380 mM NaCl, 10% glycerol, 0.4mM EDTA). After 1 hour the reaction was stopped and the proteins denatured by adding 2X SDS protein loading buffer and heated. The reaction was analyzed by SDS-PAGE, which separates the reaction mixture based on molecular

size as the proteins migrate at different rates due to impedance by polymerized polyacrylamide, toward the positive pole in an electric gradient. The protein bands were visualized by staining the polyacrylamide gel with gelCodeBlue (Thermo Scientific) (Figure 2).

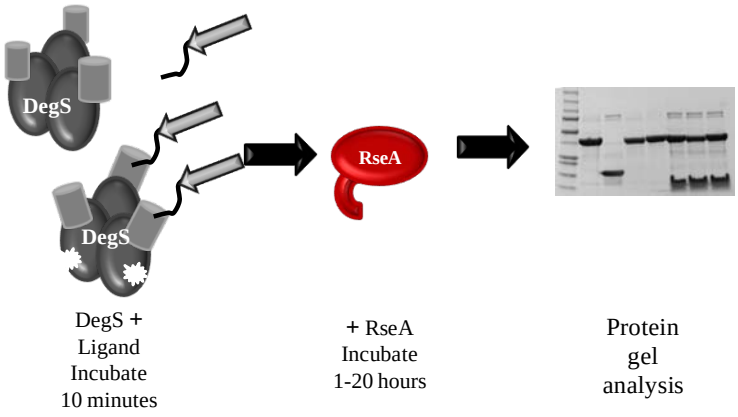


Figure 2. SDS-PAGE Analysis.
Purified, soluble DegS is incubated with activating peptide for 10 minutes at 37 °C. The substrate, the periplasmic domain of RseA, is added and the components are incubated for one to three hours. The reaction mixture is then denatured, analyzed by gel electrophoresis, and visualized by gelCodeBlue.

All the proteins appeared at least 90% pure by SDS-PAGE analysis, although variant DYF EP1.1-fusion and DYF EP1.2-fusion had double bands, indicating a potential degradation product or the protein may have two different conformations that interact with the cross-linked gel differently despite being denatured and the YYF EP1.1-fusion concentration was initially underestimated (Figure 3, Panel L). All of the proteins exhibited activity without the activating ligand, except for DegS and N285I. M319K had lower basal activity than the other four selected variants. YYF EP1.2 had the lowest basal activity amongst the four selected variants (Figure 3, Panel R).

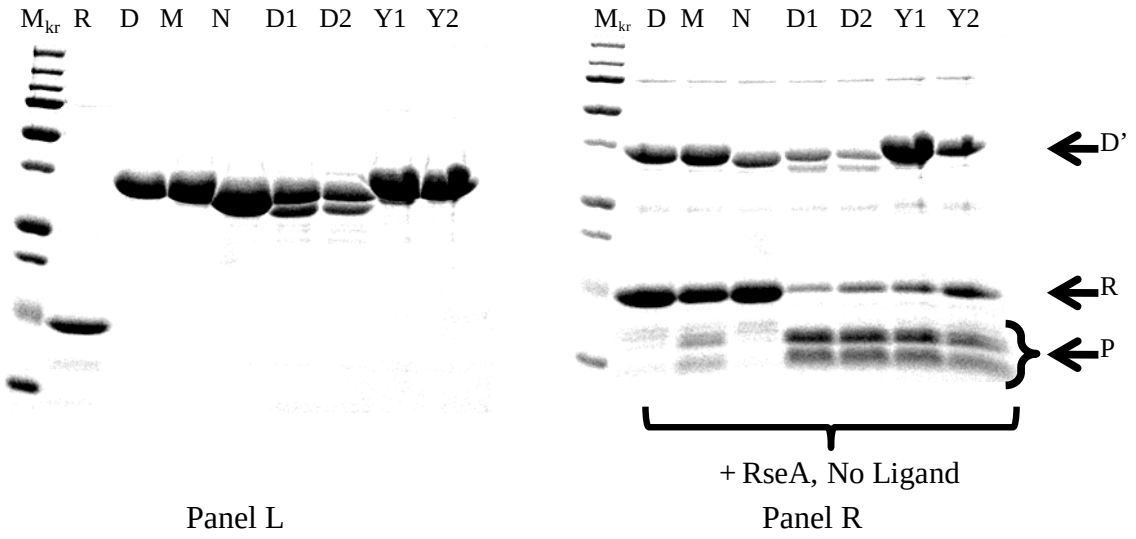


Figure 3. Control Protein Gels: Panel L) Protein and Panel R) Protein and substrate, no ligand
M_{kr} = marker, R= RseA periplasmic domain, D=soluble DegS, M= M319K, N= N285I, D1 = DYF EP1.1, D2 = DYF EP1.2, Y1 = YYF EP1.1, Y2 = YYF EP1.2. Arrow D' points to soluble DegS and variant bands (33 kDa); Arrow R points to RseA periplasmic domain (13 kDa); Arrow P points to RseA cleavage products (*C*-terminal fragments is 8 kDa, *N*-terminal fragment is 5 kDa).

The proteins were then evaluated with the two 10-mer peptide ligand activators: the Class II activator, DNRDGVNYYF (YYF), and the Class III activator, DNRDGVNDYF (DYF). The same reaction mix for the controls was used but with Class III activator, *C*-terminus DYF to evaluate protease activity related to the DYF activating peptide ligand. The protein band profiles resulting with DYF (Figure 4, Panel L), looked similar to the protein band profile results without any activator (Figure 3, Panel R). The protein band profiles resulting from the reaction components with wild-type Class II activator, *C*-terminus YYF (Figure 4, Panel R), showed that all the proteins were more active except for YYF EP1.2-fusion. The YYF EP1.2-fusion protein band profile looked similar in all three test reactions (1. no activator, 2. plus DYF, and 3. plus YYF). N285I and the wild-type DegS were the most active with YYF activator, completely degrading the available RseA substrate.

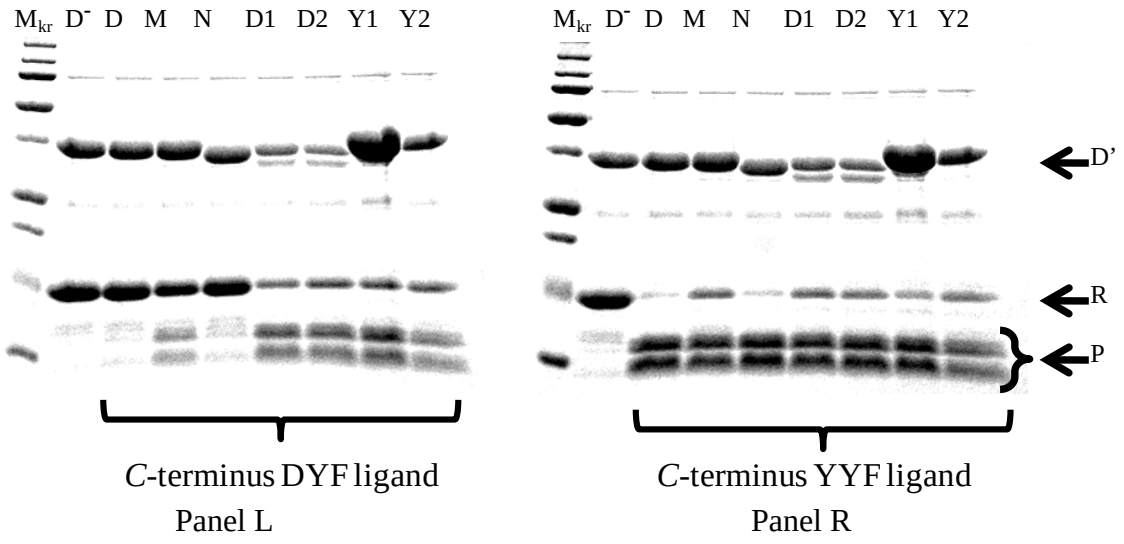


Figure 4. Protein Gels: Panel L) + C-terminus DYF ligand and Panel R) + C-terminus YYF ligand
M_{kr} = marker, R= RseA periplasmic domain, D=soluble DegS, M= M319K, N= N285I, D1 = DYF EP1.1, D2 = DYF EP1.2, Y1 = YYF EP1.1, Y2 = YYF EP1.2. Arrow D' points to soluble DegS and variant bands (33 kDa); Arrow R points to RseA periplasmic domain (13 kDa); Arrow P points to RseA cleavage products (*C*-terminal fragments is 8 kDa, *N*-terminal fragment is 5 kDa).

p-nitroanaline (*p*NA) Assay

The two differences between the SDS-PAGE and *p*NA assay are the substrate and the method of product analysis. The substrate is a commercially available synthetic substrate *N*-methoxusuccinyl-AlAlaProVal-p-nitroanalide (MeOSu-AAPV-pNA) (SigmaAldrich) (Figure 5).

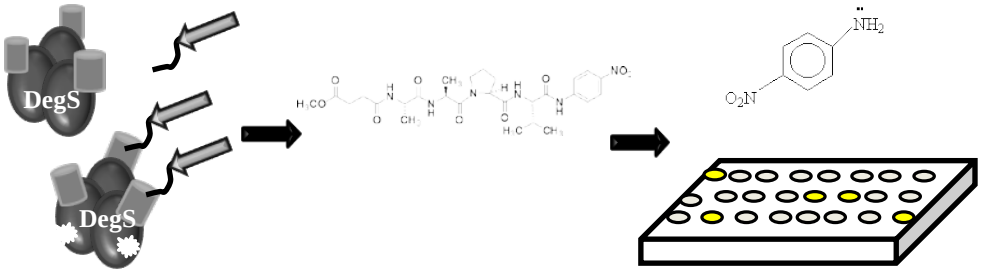


Figure 5. *p*-Nitroanaline Assay.
Purified, soluble DegS is incubated with activating peptide for 10 minutes at 37 °C. The substrate, AAVP-pna, is added and the components are incubated for one to twenty hours at 37 °C. The reaction is analyzed by spectrophotometer at Abs_{410}.

The substrate was selected after observing DegS cleave peptide substrates while evaluating experimental conditions for other potential assays. A 9mer peptide was tested with purified wild-type DegS and variant, YYF EP1.1, while investigating a potential FRET assay. The 9mer, WCAPVSLKG, was cleaved selectively with activating peptide ligand present and was not cleaved without activating peptide ligand (YYF). The reaction products were confirmed by ESI-MS performed by the University of Texas College of Pharmacy (Table 2).

Table 2: DegS and Peptide substrate ESI-MS Results. Products detected by ESI-MS after incubating with Class II peptide ligand, YYF, and wild-type DegS and a variant, YYF EP1.1-fusion, overnight at 37°C with the 9mer peptide.

Product	MW	9mer only	DegS+ 9mer +YYF	YYF EP1.1 +9mer +YYF
WCAPVSLKG*	961	+	+	+
WCAPV	576	-	+	+
SLKG	403	-	-	-

In the *p*NA assay, the reactions were analyzed in a 96 well plate format permitting higher throughput analysis than the SDS-PAGE assay. The cleavage after the valine in the substrate, MeOSu-AAPV-pNA, results in free chromogenic *p*-nitroanalide (*p*NA), which can be monitored at $Abs_{410.}$ The intensity of the emission at Abs_{410} increases proportionately to the amount of free *p*NA. The reaction was monitored continuously for the first 60 minutes and then evaluated at individual time points at 120 minutes and after overnight incubation (15 hours). The one and two hour time points indicated activity; however, the results were not statistically significant (data not included). The overnight time point data were consistent over three independent measurements and differences between each DegS fusion were distinguishable.

Initially, it was attempted to characterize the reaction kinetics with the MeOSu-AAPV-pNA substrate. Five concentrations between 0.5 and 10 mM of MeOSu-AAPV-pNA substrate were monitored over 212 minutes with three different concentrations of wild-type DegS protein (6.3, 12.5 and 25 µM of DegS). These data points were obtained in duplicate and included three controls of protein with no substrate and five controls of substrate with no protein. Unfortunately, the data are not conclusive because it does not represent a linear, steady-state rate increase of *p*NA product over the substrate concentration gradient; and thus, standard enzyme kinetics analysis by the Michealis-Menton equation does not apply. Instead, the plot of velocity (nmole of product produced per second per nmole of enzyme) versus MeOSu-AAPV-pNA substrate concentration (mM) is exponential (Figure 6), indicating non-first order kinetics, which is expected due to ligand and/or substrate binding cooperative effects amongst the trimer sub-units. K_m is the substrate concentration at which half maximal enzyme velocity is reached and in some cases represents the affinity of the enzyme-substrate interaction. The 10mM substrate concentration was near the reaction solution saturation point with up

to 10% by volume solvent (dimethylformamide (DMF)) as increasing the substrate further resulted in precipitates. As a result, the kinetic analysis was discontinued and experimental results were based on relative activity normalized with wild-type DegS basal activity (activity without activating peptide ligand).

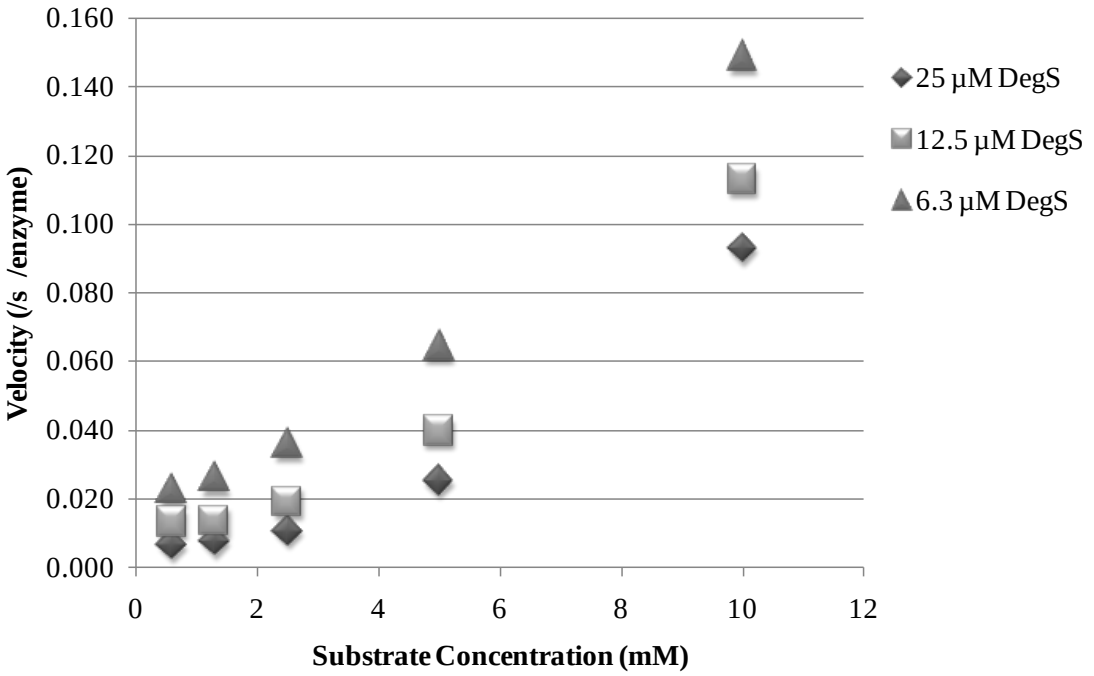


Figure 6: MeOSucc-AAPV-pNA Substrate dependence on the cleavage rate (Velocity) of the wild-type DegS enzyme. Data was analyzed for three concentrations of DegS (the diamond is 25 µM of DegS, the square is 12.5 µM of DegS and the triangle is 6.3 µM of DegS). There was not sufficient data to complete an error analysis. Since the reaction was not at steady state velocity with up to 10mM of substrate, which was near the point of substrate saturation of the solution, kinetic analysis was discontinued for this enzyme and substrate.

The *p*NA reactions each contained 6µM of DegS wild-type protein or of the DegS fusion proteins and 2mM of the MeOSucc-AAVP-pNA substrate with six to seven dilutions of YYF or DYF and one reaction containing no activating peptide ligand. The results are normalized to wild-type DegS basal activity, which is the wild-type DegS proteolytic activity without activating ligand. Variant N285I and wild-type DegS have

similar activity profiles with the highest activity at nearly 6-fold over baseline when incubated with varying concentrations of YYF. Variant M319K only increased 2-fold over baseline activity at the highest concentration of YYF (Figure 7). Both of the DYF variants do not appear to have allosteric activity, albeit there was larger variation in the data. The two DYF variants are more active than baseline, regardless of the YYF concentration (Figure 8). YYF EP1.1 is the only variant that has increasing activity with increasing ligand concentration, although its basal activity is 3-fold higher than wild-type DegS basal activity. Variant YYF EP1.2 activity remained 2-fold more active than baseline for all concentrations of ligand and without ligand (Figure 9).

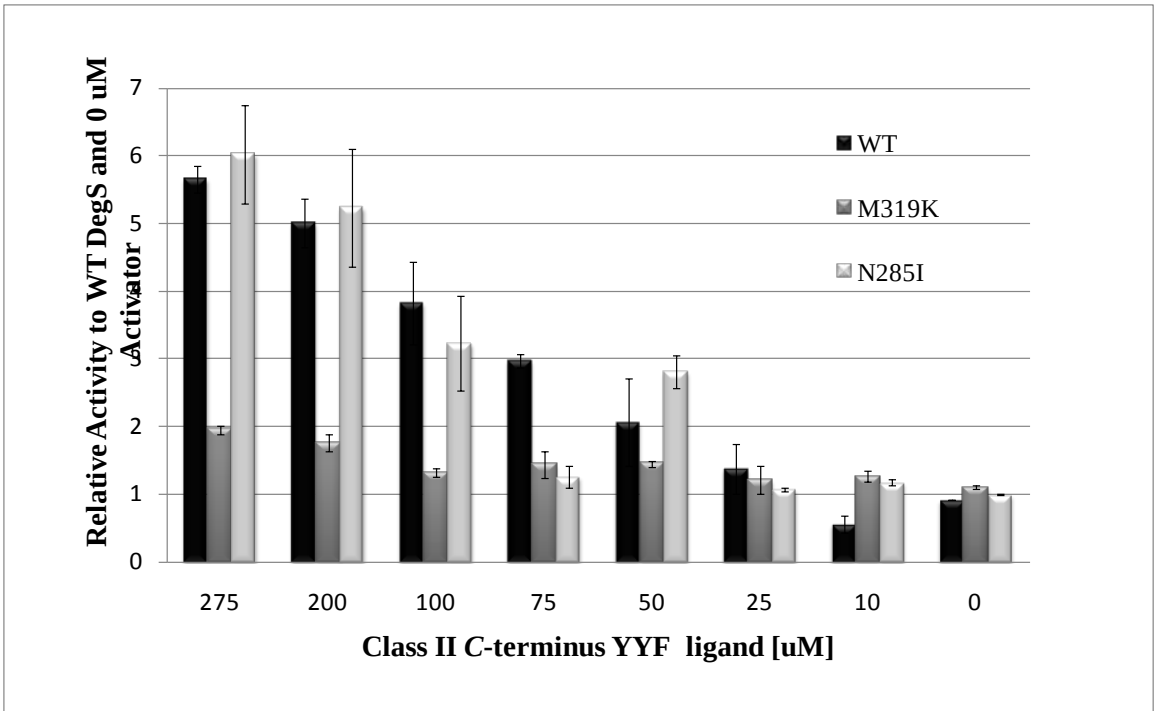


Figure 7. *p*NA assay: WT, M319K and N285I and YYF ligand. 6µM wild-type DegS, Variants M319K and N285I were incubated with a gradient of concentrations of YYF and 2mM of MeOSucc-AAPV-pNA substrate for a 15 hours at 30 °C. The data represents the average of three measurements and the error bars are the respective standard deviation.

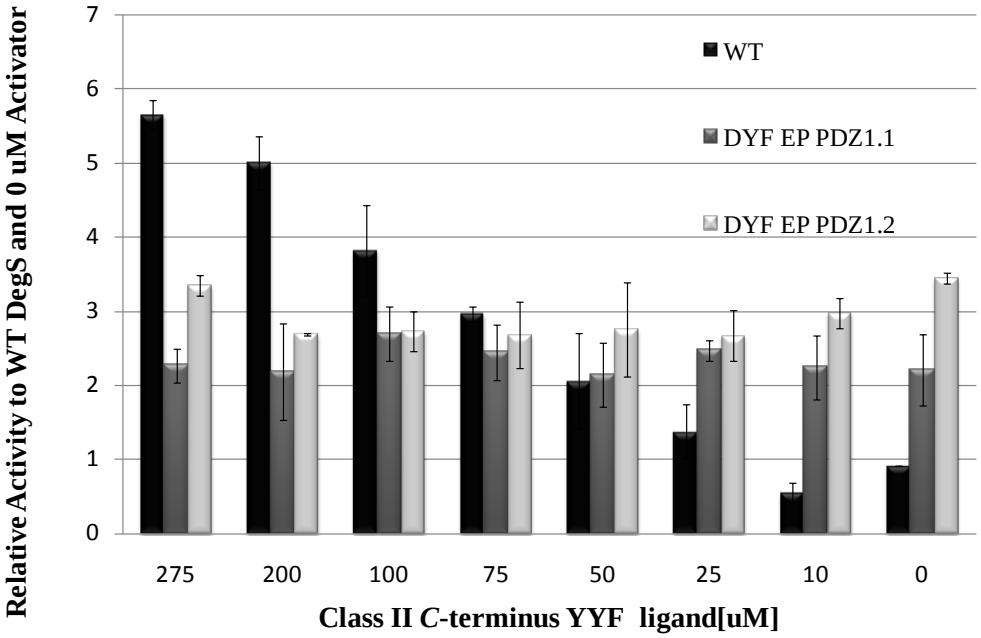


Figure 8. *p*NA assay: WT, DYF EP1.1 and DYF EP1.2 and YYF ligand. 6µM wild-type DegS and Variants DYF EP1.(1&2) were incubated with a gradient of concentrations of Class II YYF and 2mM of MeOSucc-AAPV-pNA substrate for a 15 hour incubation time. The data represents the average of three measurements and the error bars are the respective standard deviation.

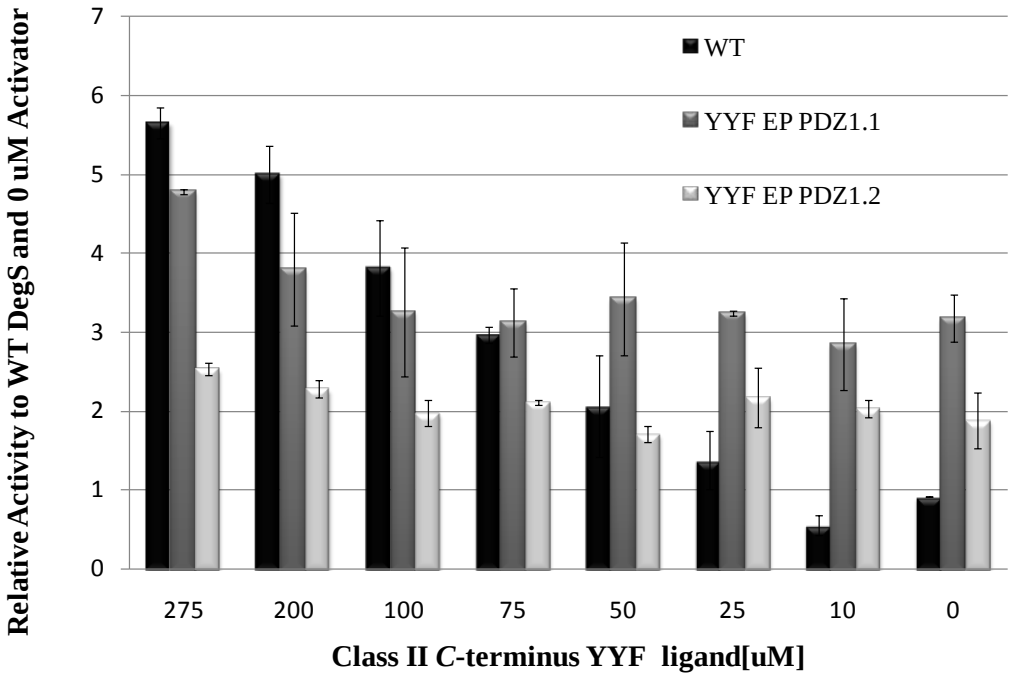


Figure 9. *p*NA assay: WT, YYF EP1.1 and YYF EP1.2 and YYF ligand. 6µM wild-type DegS and Variants YYF EP1.(1&2) were incubated with a gradient of concentrations of YYF and 2mM of MeOSucc-AAPV-pNA substrate for a 15 hour incubation time. The data represents the average of three measurements and the error bars are the respective standard deviation.

Variant M319K activity increases over baseline at the highest concentration of DYF activating peptide ligand, however, it is within the data standard deviation. Wild-type DegS and variant N285I have no protease activation above basal activity following the overnight incubation with the DYF activating peptide ligand (Figure 10). When incubated with the DYF ligand, the DYF variants appear to have no overall increase in activity over baseline activity. Due to the data variation, a more definitive statement cannot be made regarding the DYF variants and the DYF activating peptide ligand (Figure 11). The YYF EP1.1 variant reflects an increase in activity with increasing concentration of the DYF ligand (Figure 12); however to a lesser extent than it did with

the YYF activating peptide ligand (Figure 9). Variant YYF EP1.2 activity remained above wild-type DegS basal activity for all concentrations of DYF peptide ligand and no ligand; however, there was also data variability in the YYF EP1.2 data set (Figure 12).

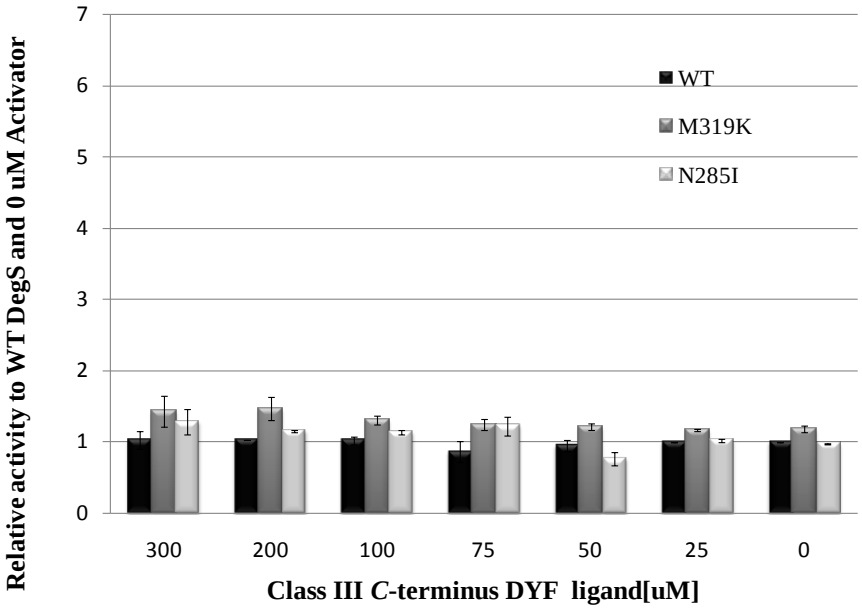


Figure 10. *p*NA assay: WT, M319K and N285I and DYF ligand. 6µM wild-type DegS, Variants M319K and N285I were incubated with a gradient of concentrations of Class III DYF and 2mM of MeOSucc-AAPV-pNA substrate for a 15 hour incubation time. The data represents the average of three measurements and the error bars are the respective standard deviation.

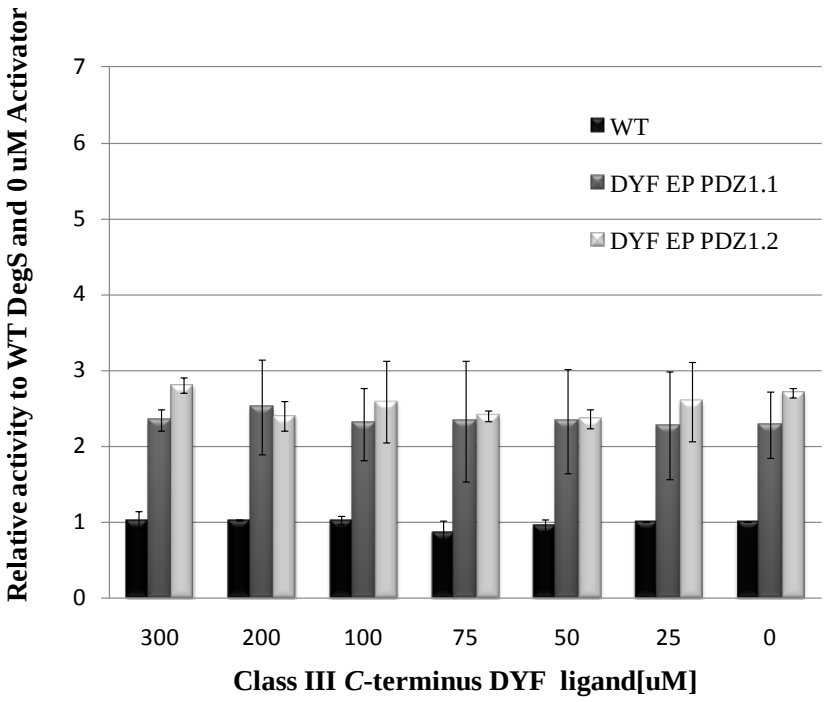


Figure 11. *p*NA assay: WT, DYF EP1.1 and DYF EP1.2 and DYF ligand. 6µM wild-type DegS and Variants DYF EP1.(1&2) were incubated with a gradient of concentrations of Class III DYF and 2mM of MeOSucc-AAPV-pNA substrate for a 15 hour incubation time. The data represents the average of three measurements and the error bars are the respective standard deviation.

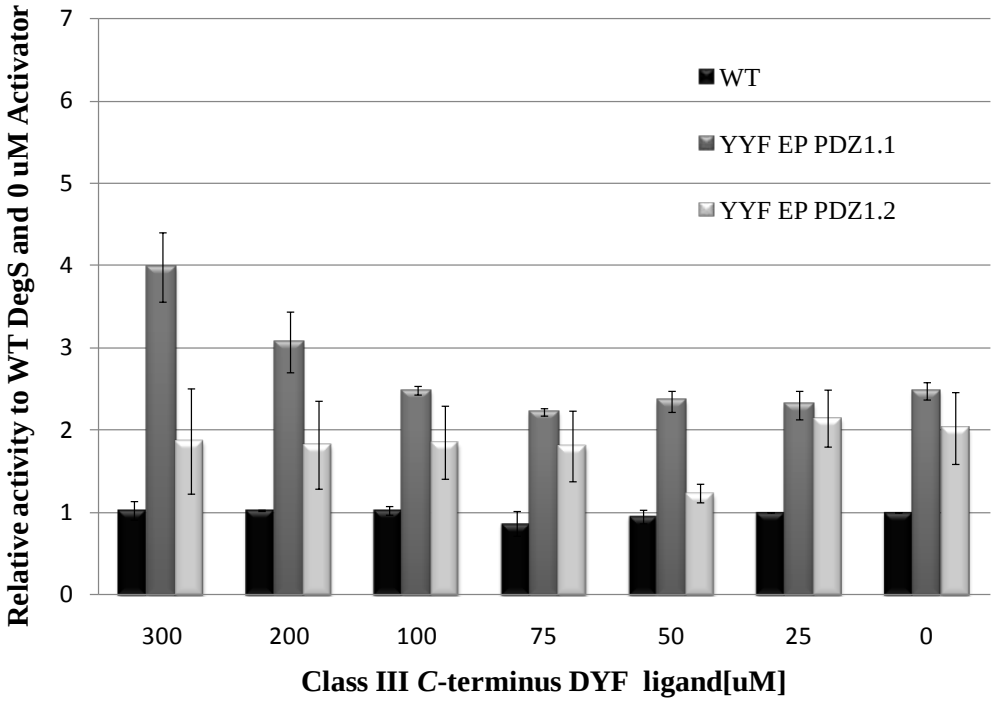


Figure 12. *p*NA assay: WT, YYF EP1.1 and YYF EP1.2 and DYF ligand. 6µM wild-type DegS and Variants YYF EP1.(1&2) were incubated with a gradient of concentrations of Class III DYF and 2mM of MeOSucc-AAPV-pNA substrate for a 15 hour incubation time. The data represents the average of three measurements and the error bars are the respective standard deviation.

The *p*NA data for each enzyme and ligand concentration profile provides a basis for evaluating each enzyme's allosteric property; however, to more concisely compare the allosteric properties of each enzyme evaluated in this investigation the enzyme activity at maximum ligand concentration is divided by the enzyme activity without ligand. If the activity is the same with and without ligand, the enzyme is not allosteric and the value of this ratio equals 1. The benchmark for the ratio of activity with maximal ligand concentration to the activity with no ligand is the DegS wild-type enzyme.

Wild-type DegS and the N285I construct both had similar activity ratios, 5.6±0.2 and 6.0±0.7 for the YYF activating peptide ligand. The activity ratio for M319K is 1.8±0.1 for the YYF ligand. YYF EP1.1 activity ratios for YYF and DYF peptide ligand

are 2.6±0.3 and 2.5±0.5 respectively. The ratios for the other DegS variant enzymes were either not statistically significant due to data variability and propagating that error through the calculations or indicated no allostery with ratio values of 1.0 within experimental the error (see Figure 13).

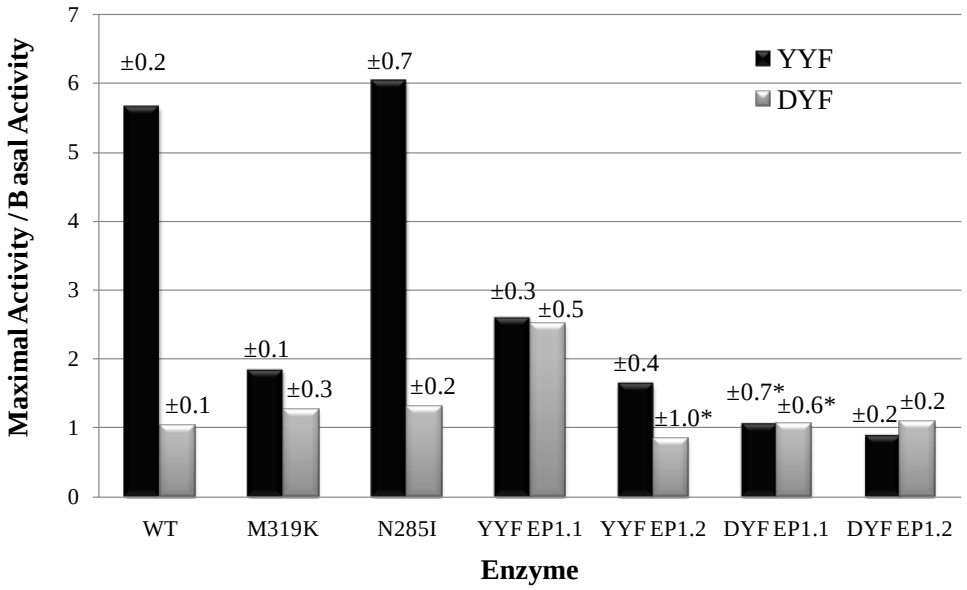


Figure 13. *p*NA assay Allosteric Activity Comparison. The maximal activity for each enzyme at the highest activating ligand concentration is divided by the basal activity for each respective enzyme to compare allostery of each enzyme variant. WT is the wild-type DegS enzyme. Above each bar is the propagated error value. The error values marked with * indicate the data is not significant.

DISCUSSION

The objective was to determine if the protein fusions comprised of the DegS wild-type protease domains and the four previously selected variant PDZ domains, are allosteric. Additionally, two enzyme variants comprised of PDZ domain constructs with single consensus mutations were used to discern the impact of mutations in the selected variants. A conventional SDS-PAGE assay was used to qualitatively determine if the protein fusions were allosteric. Additionally, a *p*NA assay was used to further probe

allostery with a gradient of activating ligand concentrations. The *p*NA method was not used for evaluating DegS proteolytic activity until recently by the Clausen group, while this dissertation work was in progress [185]. The Clausen group synthesized a preferred DegS enzyme substrate, H-VFNTLPMMGKASPV-*p*NA, which consists of fourteen amino acids equivalent to the RseA *N*-terminal cleavage sequence. I used a commercially available substrate, MeOSucc-AAPV-*p*NA, which is used as a substrate for human and mouse neutrophil elastase (leukocyte elasatase) and neutrophil proteinase 3 (PR-3, a myeloblastin) (Sigma Aldrich).

Comparison of SDS-PAGE and pNA Assay Results

It is expected that the two methods should corroborate each other qualitatively, if they are both valid approaches. Both assays demonstrated a successful positive control with wild-type DegS exhibiting allosteric activation with the wild-type Class II activating peptide ligand (YYF) and no allosteric activation with the Class III activating peptide ligand (DYF). The single point mutation variant, N285I, had results similar to wild-type DegS by both assay methods. The SDS-PAGE assay indicated M319K basal activity was above wild-type DegS and N285I, but below the other variants selected from a random mutagenic library. The results from both assays are consistent for the controls and the two protein constructs, M319K and N285I.

The SDS-PAGE assay indicated that the DegS protease fusions with the selected PDZ domains did not have significant allosteric activity with either ligand apart from YYF EP1.1, which had increased protease activity with activating peptide ligand YYF. The *p*NA assay results were consistent with the SDS-PAGE assay, except the *p*NA assay also indicated there was an increase in YYF EP1.1 protease activity for both increasing YYF and DYF activating peptide ligand concentrations. The two assays remain qualitatively consistent because the SDS-PAGE assay was used to examine only two

activating peptide ligand concentrations (30 μM (data not shown) and 60 μM) plus basal activity. In the SDS-PAGE assay range of activating ligand concentrations, the YYF EP1.1 activity level also did not change significantly. In fact, the SDS-PAGE assay was more sensitive to ligand concentration because it detected the difference between YYF EP1.1 basal activity and activity with 60 μM of YYF peptide ligand. This sensitivity may result from the SDS-PAGE assay using the preferred DegS protease substrate, RseA, which has a published V_{max} of 1.1±0.2 $s^{-1}enzyme^{-1}$ and K_m of 750±120 μM with the same conditions used in this investigation [186]. The K_m of the MeOSucc-AAPV-*p*NA substrate appears to be much higher based on an abbreviated kinetic analysis and thus the *p*NA substrate used in this study is much less favored by DegS than the RseA substrate. Regardless, the *p*NA and the SDS-PAGE assay are consistent, validating the assay approaches; and the *p*NA assay provides additional data to probe allostery due to its higher sample throughput since it uses a 96 well plate analytical format versus a 10 to 12 lane polyacrylamide gel format.

Evaluation of Allostery

If wild-type DegS allosteric activation with a wild-type activating peptide ligand is the benchmark for allostery in this investigation, then none of the selected PDZ domain variants fused to the DegS protease domain were allosteric. If allostery is defined as increasing protease activity with activating peptide ligand by at least 2-fold over basal activity, then YYF EP1.1 is allosteric for the Class II and Class III activating peptide ligands (refer to Figure 13).

The other three variant fusions with PDZ domains selected by library evolution were constitutively more active than wild-type DegS basal activity, but less active than activated wild-type DegS. Interestingly, the DegS PDZ domain with the highest binding affinity (K_D) for both the peptide ligands, DYF EP1.2, did not transform the DegS

protease domain into an allosteric protein with a new or faster "switch". Instead, fusing DYF EP1.2 to the wild-type protease domain negated the allosteric interaction between the binding and the protease modules. In fact, the three variants that had PDZ domains with a binding affinity (K_D) for the DYF peptide ligand of less than 1 μM did not exhibit allostery for either activating peptide ligand.

The variant construct N285I, retained its allosteric property with activating peptide ligand YYF. However, despite having a K_D of 27.95 ± 5.10 μM with the DYF peptide ligand, it was not activated with 300 μM of DYF. The variant construct M319K was not activated significantly with either YYF or DYF increasing ligand concentrations even though M319K also had higher binding affinities for these ligands than wild-type DegS.

Increasing the protein binding module's affinity for an activating peptide ligand did not correlate to increasing the activation of the protease domain. Only one protein variant, YYF EP1.1, was allosterically activated by a novel activating peptide ligand, DNRDGNVDYF, despite being engineered to bind with higher affinity for the wild-type ligand. Thus, the YYF EP1.1 allosteric activation "switch" became less selective when engineered for higher binding affinity; but, notably, a degree of allosteric activation was retained.

Structural analysis

An analysis of the mutations in the context of the protein tertiary structure can provide insight regarding amino acids to target or avoid for future DegS protein engineering attempts. First, an examination of wild-type DegS structures with ligand bound and unbound reveal three notable conformation changes between the active and inactive protease conformation: 1) the protease domain Loop 3 (L3) conformation shift away from the PDZ domain; 2) the stabilization of the catalytic site by protease domain

Loop 2 (L2) and His198, Leu164, and Gln166; and 3) the formation of the S1 substrate binding pocket also stabilized by L2. L3 has several interactions with the unbound PDZ domain including two salt bridges: one formed by protease domain residue, Arg178, and PDZ domain residue, Asp320, and the other formed by protease domain residue Lys243, and PDZ domain residue, Glu324. When the PDZ domain binds the ligand, the PDZ domain αB helix (residues 314 – 324), which interacts directly with the ligand, appears to rotate and cause the salt bridges to dissociate. The dissociation of these salt bridges permits L3 to shift down, which displaces Arg178 by approximately 17 angstroms. Arg178 is then free to interact with the neighboring DegS sub-unit residues Leu164 and Gln166, which shift their conformation and in turn causes changes of polar and hydrophobic contacts between residues resulting in the His198 residue to rotate around the protein backbone and stabilize the correct conformation of the catalytic triad. Protease domain L2 is a flexible loop that disrupts and blocks the substrate binding and catalytic sites in inactive DegS. This loop becomes rigid and contributes to stabilizing the substrate and catalytic sites in active DegS (see Figure 14).

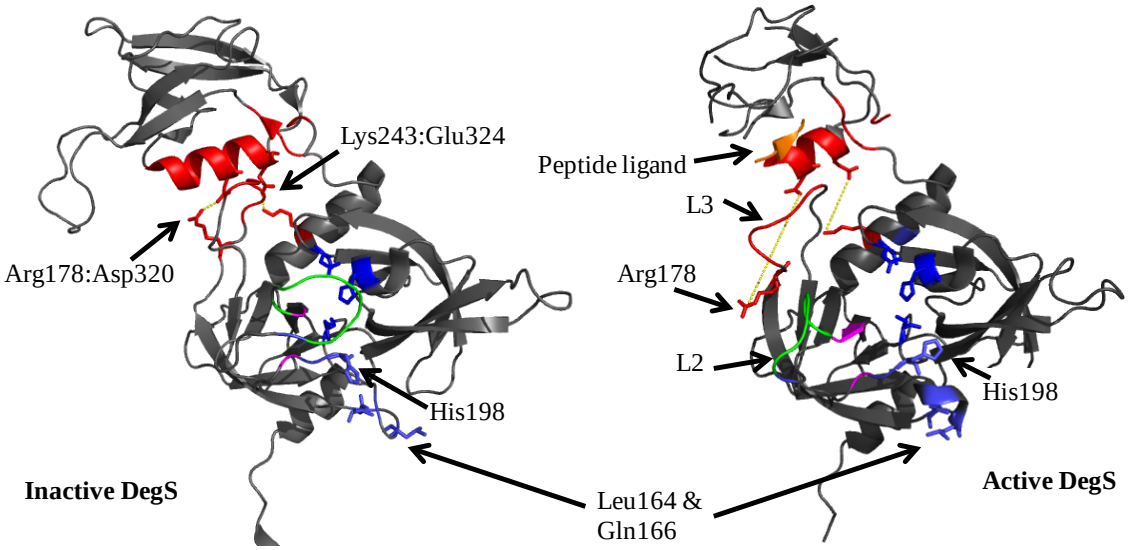


Figure 14. Inactive DegS (PDB:1TEO [184]) and Active DegS (PDB:1SOZ [184]) with the viewer looking down at the catalytic active site. Residues in the protease and PDZ domains that are within 4 angstroms of each other when DegS is inactive are red with the two salt bridges Arg178:Asp320 and Lys243:Glu324 that assist in keeping L3 proximal to the PDZ domain labeled. L3 shifts away from the PDZ domain in active DegS dissociating the Arg178 and K243 salt bridges. The bright blue residues are the catalytic triad (His96, Asp126 andSer201), which is accessible in Active DegS with the correct orientation of Ser201. His198 (light blue) rotates around the protein backbone to help form the active catalytic site. Leu164 and Gln166 (both light blue) interact with the neighboring DegS sub-unit's Arg178 to further stabilize the active catalytic site. L2 blocks access to the catalytic site in inactive DegS and then becomes more rigid in active DegS and stabilizes the S1 substrate pocket (in pink). The orange β strand is the activating peptide ligand. Structures modeled using PyMol [104]

The two constructs with the consensus mutations, N285I and M319K, are informative since N285I retained allostery for the wild-type ligand and M319K had reduced allostery for the wild-type ligand. Additionally, they both were not activated by the DYF activating peptide ligand despite having improved binding affinity for the DYF ligand over the wild-type DegS enzyme. Asn285 is distal from PDZ domain residues that interact with the protease domain and results in minimal impact to allostery. On the other hand, Met319 is located on the PDZ domain αB helix, which interacts directly with the protease domain. The reduced allostery in M319K suggests that changing Met319 to a Lys results in reducing the interactions between the PDZ and protease domains such that

L3 is less stable in the inactive DegS conformation (see Figure 15). If this were true, then the M319K protein fusion should be constitutively active. The M319K construct has low proteolytic activity suggesting that the altered interactions resulting from the mutation may have stabilized a different L3 structure from both the active and inactive conformations. The structure also does not clearly indicate why the PDZ domain alone has improved binding affinity for the DYF peptide ligand (particularly M319K, which has 9-fold better binding affinity for the DYF peptide ligand than the N285I PDZ domain) yet when fused to the protease domain does not activate the protease. It is possible the mutations altered the PDZ domain to permit it to bind to the DYF ligand without impacting the protease domain because a key element (for example, the αB helix structure) in the allosteric communication pathway between the binding event and the proteolytic event was changed. It is also possible that the DYF peptide ligand did not bind to the DegS fusion constructs as binding equilibrium dissociation constants (K_D) were not measured for the full length fusion enzymes.

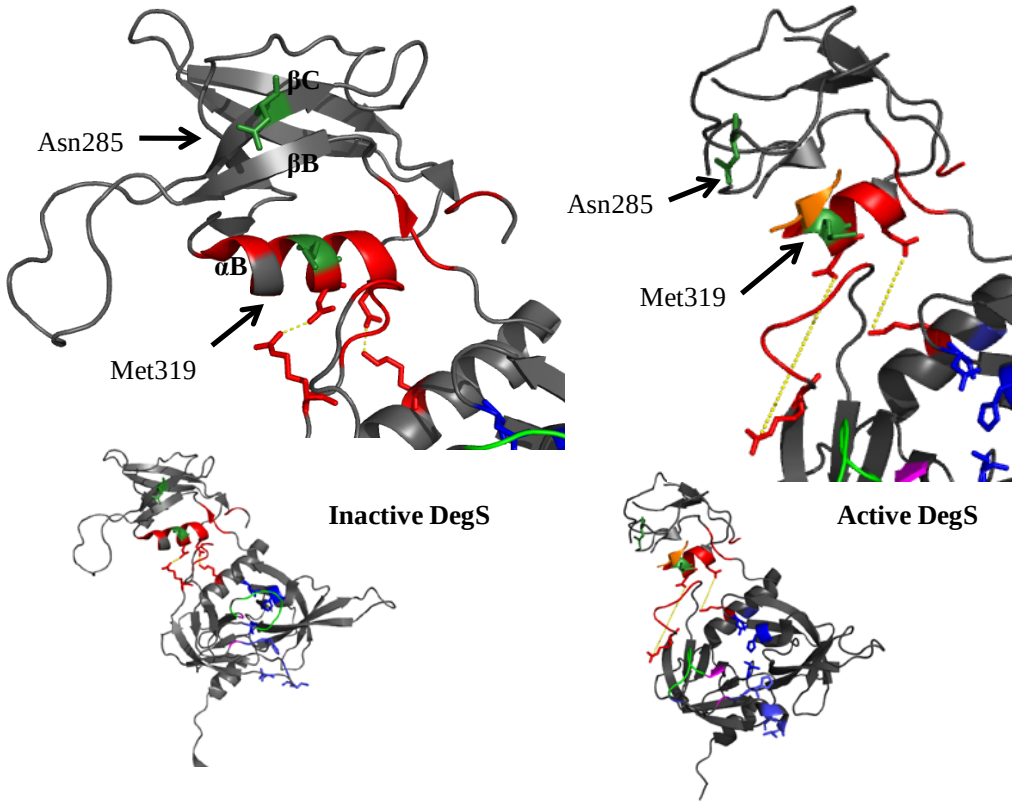


Figure 15. Inactive DegS (PDB:1TEO [184]) and Active DegS (PDB:1SOZ [184]) structures with the N285 and M319 residues labeled and highlighted in green. The entire structures are included for perspective and the PDZ domains are enlarged. The red residues potentially interact between the PDZ and protease domain in Inactive DegS. The orange β strand is the activating peptide ligand. Structures modeled using PyMol [104]

The three variants with no allosteric activity, the YYF EP1.2-fusion and the DYF EP1.1 and DYF EP1.2 –fusions, all had the consensus mutation M319K. The M319K construct contributes to their lack of allosteric activation due to its impact on the PDZ domain αB helix. However, these three variants were constitutively more active than the M319K construct though still less active than activated wild-type DegS. Since they were constitutively more active, it is expected that protease domain must prefer a more active DegS conformation. YYF EP1.2 and DYF EP1.1 have multiple mutations in the αB helix and the loop between the βB and βC strands that can further influence the interactions between the PDZ and protease domains. Specifically, Asp320 is mutated to a Gly in DYF EP1.1 which normally forms a salt bridge with L3 through Arg178.

Unfortunately, active, ligand bound DegS structures are not resolved for residues in the loop between the βB and βC strands. This loop may influence peptide binding of ligand residues beyond residue (-4). DYF EP1.2 and YYF EP1.2 also had mutations in the loop following the βC strand: Q299R and S288T, respectively. These two mutations are distal from the binding and active sites and do not have a clear direct impact (see Figure 16).

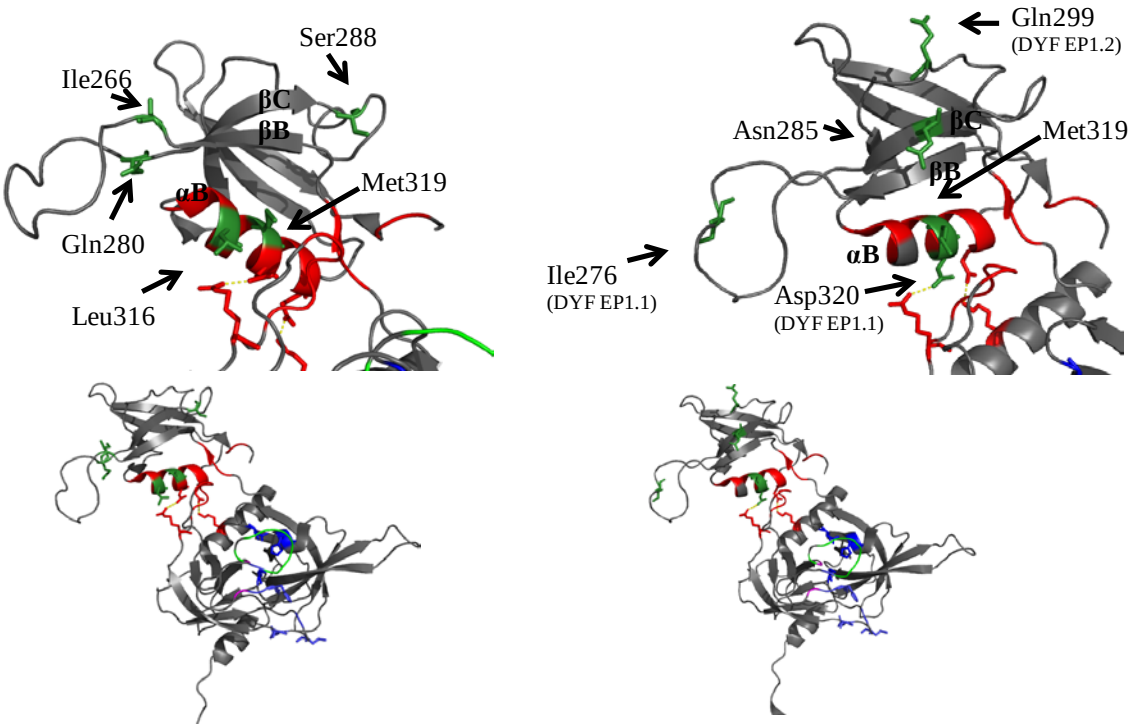


Figure 16. Inactive DegS (PDB:1TEO [184]) structures with the residues mutated in the YYF EP1.2 (left) variant and the DYF EP1.1 and DYF EP1.2 variants (right) labeled and highlighted in green. The entire structures are included for perspective and the PDZ domains are enlarged. The red residues potentially interact between the PDZ and protease domain in Inactive DegS. Structures modeled using PyMol [104].

The variant YYF EP1.1-fusion retained allosteric activation by increasing protease activity 2-fold with maximal activating YYF peptide ligand concentration albeit less than the 5-fold increase in protease activity for the wild-type DegS with maximal activating YYF ligand. Notably, the YYF EP1.1-fusion gained allosteric activation by

increasing protease activity 2-fold at the maximal novel activating DYF peptide ligand concentration. The YYF EP1.1 mutations, I276V and A312V, are located in the the loop between the βB and βC strands and just before the αB helix respectively. DYF EP1.1 also had a mutation at position Ile276 but its mutation was to a basic residue (Lys) while the YYF EP1.1 mutation to a Val is to a hydrophobic residue similar to Ile. The Ala312 mutation to a Val is a modest change from a hydrophobic residue to a larger hydrophobic residue but may have been enough of a change to impact the conformation of the neighboring αB helix; thus influencing the interactions between the PDZ and protease domains in the inactive conformation and causing higher basal activity (see Figure 17). Since the degree of allostery was the same for both activating peptide ligands, it follows that the peptide ligand residue that potentially influences protease activity is not the residue in position (-2), since this was the only residue different between the two peptide ligands. These mutations appear unexceptional, nevertheless they are impactful as they increased the YYF EP1.1-fusion variant's basal activity over wild-type, reduced the allosteric interaction between the PDZ and protease domain yet maintained a degree of allostery with wild-type activating peptide ligand and gained the same degree of allostery with a novel activating peptide ligand.

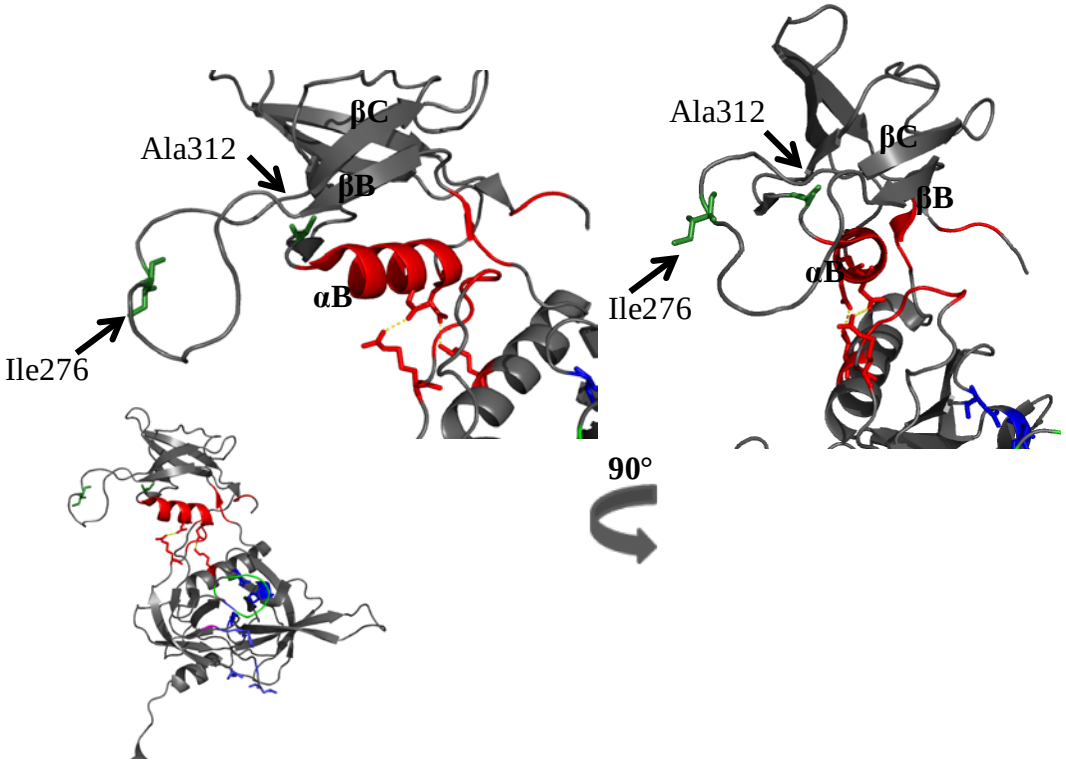


Figure 17. Inactive DegS (PDB:1TEO [184]) structures with the residues mutated in the YYF EP1.1 labeled and highlighted in green. The entire structure is included for perspective and two views of the PDZ domain are enlarged. The red residues potentially interact between the PDZ and protease domain in Inactive DegS. Structures modeled using PyMol [104].

CONCLUSION

In effect, an allosteric protein, the YYF EP1.1-fusion, with a novel activating peptide ligand was engineered using the parent allosteric protein template, DegS. The degree of activation was less than that of the wild-type DegS with wild-type activating peptide ligand; nonetheless, an increase in the YYF EP1.1-fusion activity was a direct result of an increase in activating ligand concentration. The activation of the YYF EP1.1-fusion was not specific to the activating peptide ligand residue in position (-2) since the same activation was also obtained with the wild-type activating peptide ligand. Remarkably, these changes in allostery were a result of two modest mutations in the PDZ domain of the YYF EP1.1-fusion protein.

Redesigning an allosteric protein is feasible, though complex, and the approach used in this investigation can be improved. The approach used laboratory evolution to customize the ligand binding characteristics of the parent protein's binding modules and then fuse it to the wild-type protease module. Variants with mutations prevalent in the binding module region that also interacted with the protease abolished allostery. The construction of the combinatorial library for laboratory evolution needs to focus on residues that do not interact with the protease domain directly to improve the probability of obtaining a variant that does not impact allostery. However, the PDZ domain residues appear to be considerably interconnected since the moderately successful variant YYF EP1.1's mutations were not amongst the residues known to interact directly with the protease domain or the binding ligand. Improved structures of the parent protein bound to ligand and investigating the impact of the PDZ loop between the βB and βC strands would likely provide more insight on which residues to target. The results from the N285I construct suggest that targeting residues in the βC strand may be sufficient to change ligand binding and not impact allostery, although it is not straightforward since the N285I did have improved binding to the novel activating peptide ligand but it was not allosterically activated by it.

The N285I variant demonstrates a potential drawback to engineering the modules separately of an allosteric protein whereas the characteristic obtained by engineering one module may not transfer when the modules are fused. Optimally, an allosteric protein would be engineered with the related modules to ensure allostery is retained. The full length *E. coli* DegS protein presents a challenge to engineer with a high throughput method such as flow cytometry since the protein is essential and its protease activity must be carefully regulated to have surviving cells for analysis. A parent allosteric protein template that is not essential to *E. coli* may be a better target for engineering a full length

protein. Alternatively, a 96-well plate assay with a more optimal *p*NA substrate or other chromogenic substrate may be used to monitor DegS protease activity; however, the assay would need to be optimized for cell lysates instead of purified protein. Also, given the complexity of the interconnecting relationships between amino acids in an allosteric protein a low throughput method may severely limit the probability of finding a desired variant.

If at least one module can be screened using a high-throughput method, engineering an allosteric protein by module may still be the most efficient method. An improvement, aside from targeting mutagenesis in the combinatorial library for the high throughput screen, would be to construct the cloning plasmid in such a way that the protease domain can be easily fused to the binding domain for direct follow up with an optimized 96 well plate assay that uses cell lysate to enable at least low throughput screening for allostery of the selected variants from the high throughput method.

This investigation successfully demonstrates the feasibility and the complexity of engineering an allosteric protein. Additionally, it reveals a potentially important role for the DegS PDZ domain loop between the βB and βC strands in peptide ligand binding and the allostery. The important role of the PDZ domain αB helix structure has been elucidated in previous studies.

APPENDIX I: ANALYSIS OF DEGS HOMOLOGS

DOMAIN ORGANIZATION AND PHYLOGENIC RELATIONSHIP

DegS is a member of the high temperature requirement (HtrA) proteases. HtrA proteases consist of a serine protease domain, characterized as "trypsin-like", and one or two *C*-terminal PDZ domains (see Figure 1). HtrA proteases are present in nearly all bacterial and eukarytotic species. HtrA proteases are absent from the phylum Nematoda. In a BLink (BLAST (Basic Local Alignment Search Tool) Link) search, there were 98 protein sequences found distantly related to DegS in archea, which was not expected based on a literature published in 2002, which indicated there were no HtrA protease homologs in archea [197]. However, I have not found published data confirming these hypothetical proteases in archea at this time based on searches of on-line resources provided by NCBI (National Center for Biotechnology Information). However, most searches for homologous HtrA proteins are based on DegP, the first HtrA family member discovered in *E. coli*. In mice and humans, there are four paralogs of HtrA proteins, HtrA1 through 4. HtrA 1, 3, and 4 are secreted while HtrA2 (also known as Omi) is anchored to the mitochondrial membrane.

Recently, a review was written describing the regulation of the σ^E transcription factor in different organisms [198]. DegS performs the first proteolytic cleavage of the protein (RseA) that sequesters the *E. coli* σ^E transcription factor. This review summarized proteins with a parallel function to DegS in *E. coli*. Specifically, *Pseudomonas aeruginosa* AlgW and this protein is also a homolog to DegS. The phylogenic relationship of DegS to selected homologs in other prokaryotes, the closest related hypothetical archea protease (*Methanocella paluicola*), the functionally parallel protein, AlgW, and characterized metazoan HtrA proteases is illustrated in Figure 2.

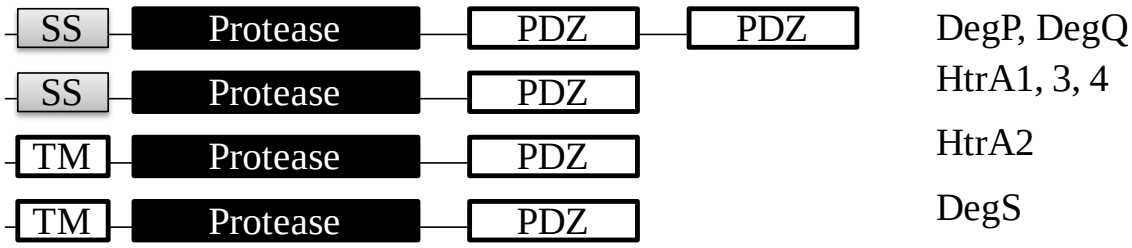


Figure 1. Domain organization of representative DegS homologs. The representation of HtrA1-4 are typical of the domain organization found in H. sapiens and M. musculus HtrA1 – 4. The HtrA2 is also representative of the D. melangaster HtrA2[199, 200].

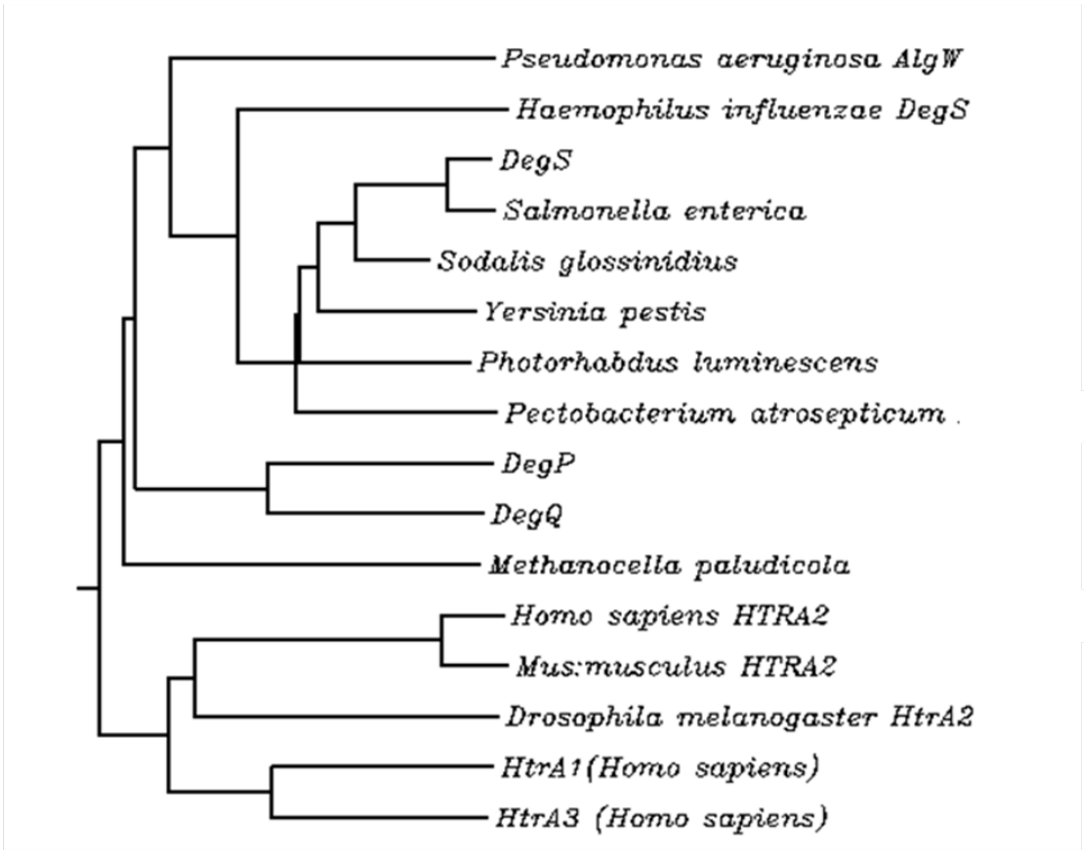


Figure 2. Phylogenic relationship of representative DegS homologs. The tree is illustrated with *TreeCon* using the neighbor joining method with branches.[201]

In *E. coli*, there are three HtrA protease paralogs: DegP, DegS, and DegQ. DegP and DegQ are closely related with significant overlapping sequence (75% sequence similarity) [202, 203]. DegP is both a protein folding chaperone and a protease in the

periplasm [204-206]. DegP's conversion from a folding chaperone to a degradative protease is currently being elucidated. DegQ is less studied and it has been suggested DegQ's function overlaps with DegP [202, 207, 208]. DegS is the only essential HtrA protease in *E.coli* and its role appears solely as a regulatory protease anchored to the inner membrane, unlike DegQ and DegP, which are periplasmic proteins[166].

In humans, HtrA2, appears to be a regulatory protease localized at the mitochondrial membrane and mutants have been implicated in Parkinson's and Alzheimer's disease[209, 210]. Other human HtrA paralogs, HtrA1 and HtrA3, have been implicated in arthritis and various cancers including melanoma, ovarian and endometrial cancer [208, 209, 211].

.

SEQUENCE ALIGNMENT

It is known that DegP substrate and protein binding specificity is more promiscuous than DegS [204, 205]. The first PDZ domain of DegP is necessary for its protease function and it binds the *C*-terminus of its protease substrate. After DegP cleaves its substrate, the substrate's new C-terminus binds to the first DegP PDZ domain and the protein is cleaved again until peptides of 12 to 17 amino acids in length remain of the substrate protein[206].

The characterized HtrA proteases indicate the PDZ domain allosterically regulates the protease domain. DegP can bind to *C*-terminus ligands that also bind to the DegS PDZ domain [204]. The human HtrA2/Omi PDZ domain binds to the *C*-terminus of ligands with similar characteristics to the DegS PDZ domain in addition to internal protein regions that are still being elucidated[209]. In general, there is a relationship

between the two characteristic domains; however, the evolved functional roles and thus the substrate specificity and selectivity of the proteins differ.

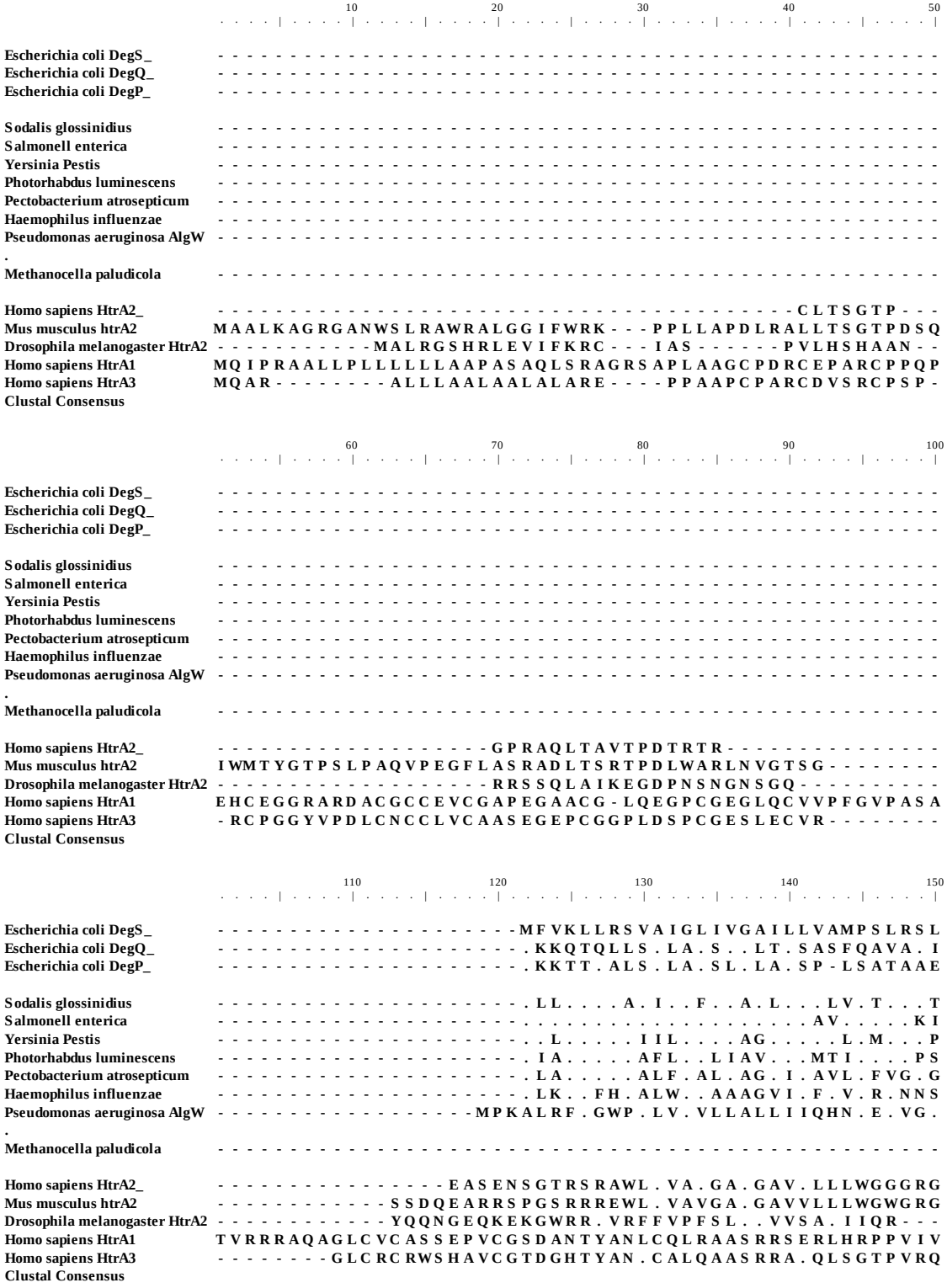
The objective was to determine which regions of the DegS homologs were conserved between species. These conserved regions potentially represent amino acids that link the PDZ protein binding domain to the serine protease domain. It is hypothesized that the protein regions with more variability between species evolved with their substrate and have more impact towards substrate binding versus the conserved allosteric communication network between the protein binding and protease domains.

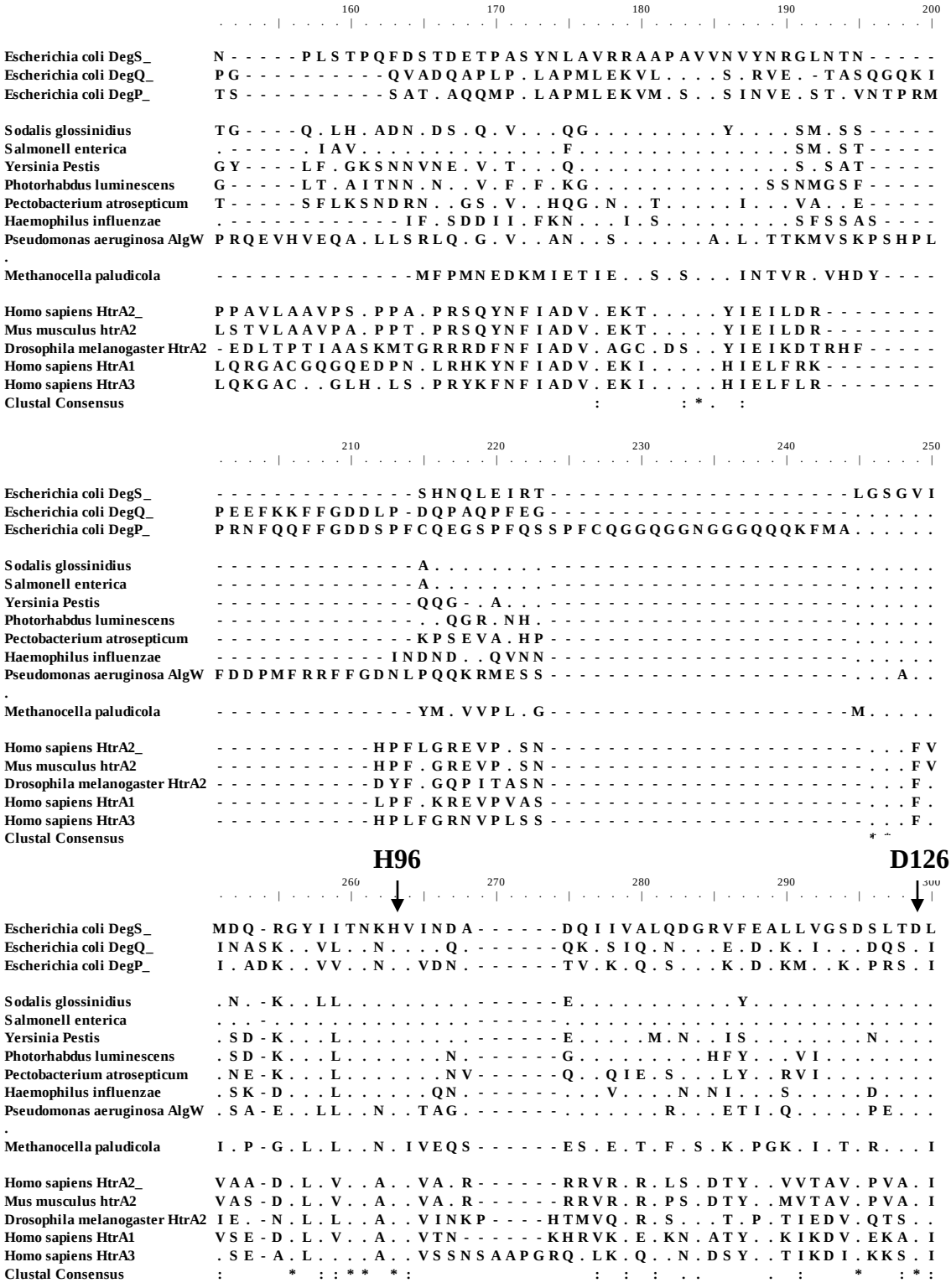
Using the NCBI's Protein Cluster program, eighty seven DegS homologs in the order *Enterobacteriales* were found. Five unique sequences of these eighty-seven closely related DegS homologs were obtained from the NCBI protein database (*Sodalis glossinidius, Salmonell enteric, Yersinia pestis, Photorhabdus luminescens, Pectobacterium atrosepticum*). Other sequences obtained for sequence alignment analysis were homologs that had been characterized as serine proteases and confirmed to function and include the homolog from *Haemophilus influenza, P. aeruginosa*, AlgW, and the acrchea homolog *M. paludicola.* The *E. coli* HtrA proteins DegP and DegQ and the human HtrA proteins 1 through3 were included along with the characterized HtrA2 proteases of *M. musculus* and *D. melangaster*. These homologs were aligned with DegS using the multiple sequence alignment program, Clustal W[201]. Figure 3 contains the alignments.

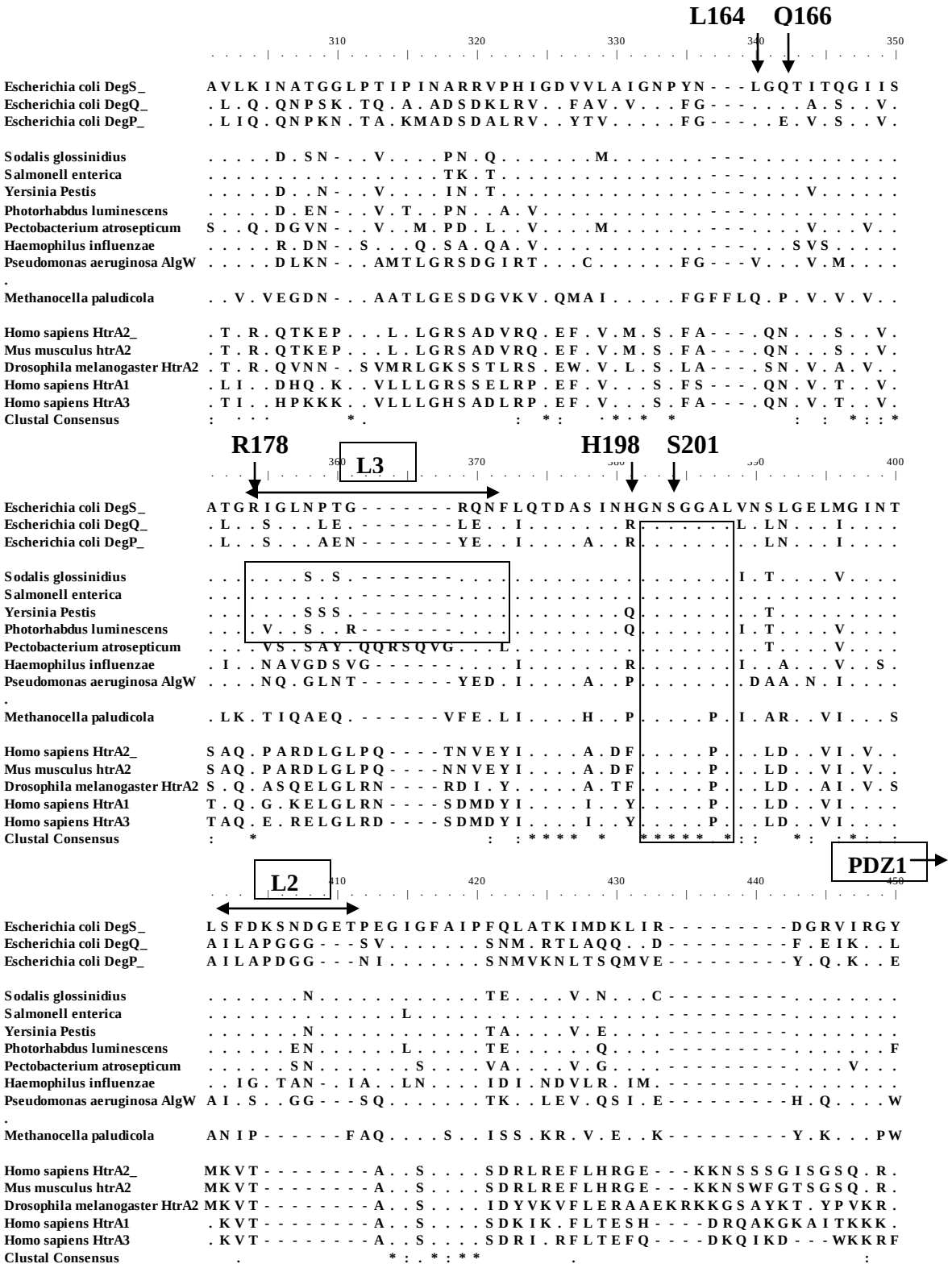
DegP consists of an extra loop after position DegP80 (located at position 220 in Figure 3) that is unique from the other HtrA proteases. This loop extension is important for DegP oligomerization [206]. Other HtrA proteases are typically trimers, including DegS and the others characterized and included in this analysis (DegQ is not characterized). The metazaon HtrA PDZ domains contain an extra α helix between the

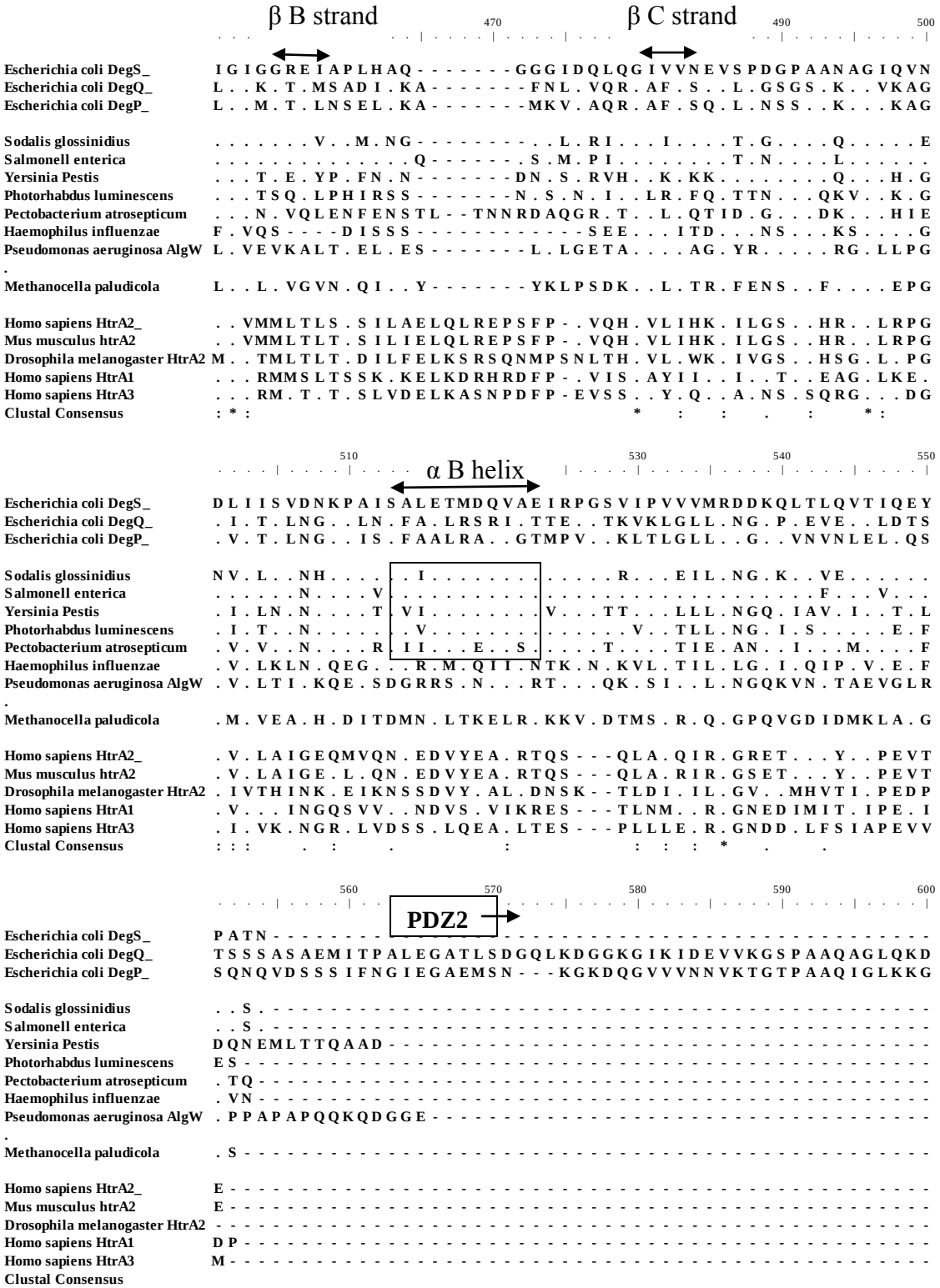
βB and βC strands, which is located at position 465 in Figure 3 and is distinguishable from the other homologs.

The catalytic sites, notated as H96, D126 and S201 (per DegS sequence numbering), are conserved across all the homologs as expected. The five DegS homologs from *Enterobacteriales* are notably conserved in the PDZ domain αB helix and the protease domain Loop 3 (both highlighted in Figure 3). These two regions appear important for allosteric activation of the protease domain as indicated by this research. There is more variability in the amino acids between the PDZ domain βB and βC strands, which indicate these amino acids may have a more significant role in ligand specificity rather than allosteric communication. Residues found important in DegS's allosteric communication network, R178 and L164, were conserved across all the homologs as well. This analysis is background for constructing targeted mutagenic libraries for the DegS protein with the goal to maintain the allosteric relationship between the protein binding domain and protease domain.


10 20 30 40 50
Escherichia coli DegS_
Escherichia coli DegQ_
Escherichia coli DegP_
Sodalis glossinidius
Salmonell enterica
Yersinia Pestis
Photorhabdus luminescens
Pectobacterium atrosepticum
Haemophilus influenzae
Pseudomonas aeruginosa AlgW
Methanocella paludicola
Homo sapiens HtrA2_
Mus musculus htrA2
Drosophila melanogaster HtrA2
Homo sapiens HtrA1
Homo sapiens HtrA3
Clustal Consensus
60 70 80 90 100
Escherichia coli DegS_
Escherichia coli DegQ_
Escherichia coli DegP_
Sodalis glossinidius
Salmonell enterica
Yersinia Pestis
Photorhabdus luminescens
Pectobacterium atrosepticum
Haemophilus influenzae
Pseudomonas aeruginosa AlgW
Methanocella paludicola
Homo sapiens HtrA2_
Mus musculus htrA2
Drosophila melanogaster HtrA2
Homo sapiens HtrA1
Homo sapiens HtrA3
Clustal Consensus
110 120 130 140 150
Escherichia coli DegS_
Escherichia coli DegQ_
Escherichia coli DegP_
Sodalis glossinidius
Salmonell enterica
Yersinia Pestis
Photorhabdus luminescens
Pectobacterium atrosepticum
Haemophilus influenzae
Pseudomonas aeruginosa AlgW
Methanocella paludicola
Homo sapiens HtrA2_
Mus musculus htrA2
Drosophila melanogaster HtrA2
Homo sapiens HtrA1
Homo sapiens HtrA3
Clustal Consensus


160
170
180
190
200
Escherichia coli DegS_
Escherichia coli DegQ_
Escherichia coli DegP_
Sodalis glossinidius
Salmonell enterica
Yersinia Pestis
Photorhabdus luminescens
Pectobacterium atrosepticum
Haemophilus influenzae
Pseudomonas aeruginosa AlgW
Methanocella paludicola
Homo sapiens HtrA2_
Mus musculus htrA2
Drosophila melanogaster HtrA2
Homo sapiens HtrA1
Homo sapiens HtrA3
Clustal Consensus
210
220
230
240
250
H96
D126
260
270
280
290
300


L164
Q166
R178
L3
H198
S201
PDZ1
L2
Escherichia coli DegS_
Escherichia coli DegQ_
Escherichia coli DegP_
Sodalis glossinidius
Salmonell enterica
Yersinia Pestis
Photorhabdus luminescens
Pectobacterium atrosepticum
Haemophilus influenzae
Pseudomonas aeruginosa AlgW
Methanocella paludicola
Homo sapiens HtrA2_
Mus musculus htrA2
Drosophila melanogaster HtrA2
Homo sapiens HtrA1
Homo sapiens HtrA3
Clustal Consensus


β B strand
β C strand
α B helix
PDZ2
Escherichia coli DegS_
Escherichia coli DegQ_
Escherichia coli DegP_
Sodalis glossinidius
Salmonell enterica
Yersinia Pestis
Photorhabdus luminescens
Pectobacterium atrosepticum
Haemophilus influenzae
Pseudomonas aeruginosa AlgW
Methanocella paludicola
Homo sapiens HtrA2_
Mus musculus htrA2
Drosophila melanogaster HtrA2
Homo sapiens HtrA1
Homo sapiens HtrA3
Clustal Consensus

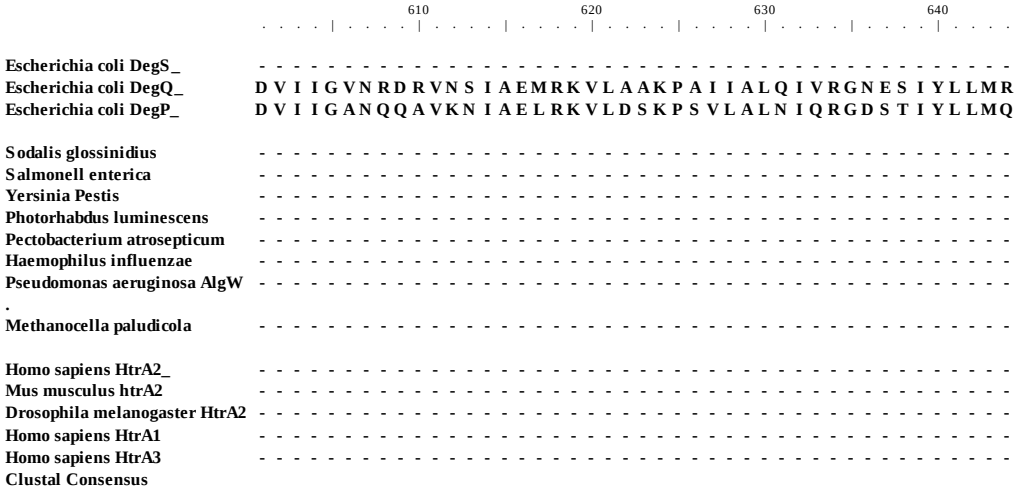


Figure 3. Sequence Alignment of DegS homologs. The sequence alignment was performed using Clustal W [201].

REFERENCES

1. Arnold, F.H. (1998). Design by directed evolution. Acc. Chem. Res *31*, 125-131.
2. Maynard, J.A., Maassen, C.B.M., Leppla, S.H., Brasky, K., Patterson, J.L., Iverson, B.L., and Georgiou, G. (2002). Protection against anthrax toxin by recombinant antibody fragments correlates with antigen affinity. Nature biotechnology *20*, 597-601.
3. Boder, E.T., Midelfort, K.S., and Wittrup, K.D. (2000). Directed evolution of antibody fragments with monovalent femtomolar antigen-binding affinity. Proceedings of the National Academy of Sciences of the United States of America *97*, 10701.
4. Carter, P.J. (2006). Potent antibody therapeutics by design. Nature Reviews Immunology *6*, 343-357.
5. Jeong, K.J., Seo, M.J., Iverson, B.L., and Georgiou, G. (2007). APEx 2-hybrid, a quantitative protein–protein interaction assay for antibody discovery and engineering. Proceedings of the National Academy of Sciences *104*, 8247.
6. Michaud, G.A., Salcius, M., Zhou, F., Bangham, R., Bonin, J., Guo, H., Snyder, M., Predki, P.F., and Schweitzer, B.I. (2003). Analyzing antibody specificity with whole proteome microarrays. Nature biotechnology *21*, 1509-1512.
7. Wark, K.L., and Hudson, P.J. (2006). Latest technologies for the enhancement of antibody affinity. Advanced drug delivery reviews *58*, 657-670.
8. Xu, L., Aha, P., Gu, K., Kuimelis, R.G., Kurz, M., Lam, T., Lim, A.C., Liu, H., Lohse, P.A., and Sun, L. (2002). Directed evolution of high-affinity antibody mimics using mRNA display. Chemistry & biology *9*, 933-942.
9. Aharoni, A., Amitai, G., Bernath, K., Magdassi, S., and Tawfik, D.S. (2005). High-throughput screening of enzyme libraries: thiolactonases evolved by fluorescence-activated sorting of single cells in emulsion compartments. Chemistry & biology *12*, 1281-1289.
10. Aharoni, A., Griffiths, A.D., and Tawfik, D.S. (2005). High-throughput screens and selections of enzyme-encoding genes. Current Opinion in Chemical Biology *9*, 210-216.
11. Arnold, F.H., and Georgiou, G. (2003). Directed enzyme evolution: screening and selection methods, (Springer).
12. Bloom, J.D., Meyer, M.M., Meinhold, P., Otey, C.R., MacMillan, D., and Arnold, F.H. (2005). Evolving strategies for enzyme engineering. Current opinion in structural biology *15*, 447-452.
13. Farinas, E.T., Bulter, T., and Arnold, F.H. (2001). Directed enzyme evolution. Current opinion in biotechnology *12*, 545-551.
14. Fernandez-Gacio, A., Uguen, M., and Fastrez, J. (2003). Phage display as a tool for the directed evolution of enzymes. Trends in Biotechnology *21*, 408-414.

15. Griswold, K.E., Kawarasaki, Y., Ghoneim, N., Benkovic, S.J., Iverson, B.L., and Georgiou, G. (2005). Evolution of highly active enzymes by homology-independent recombination. Proceedings of the National Academy of Sciences of the United States of America *102*, 10082.
16. Tao, H., and Cornish, V.W. (2002). Milestones in directed enzyme evolution. Current Opinion in Chemical Biology *6*, 858-864.
17. Graf, L., Craik, C.S., Patthy, A., Roczniak, S., Fletterick, R.J., and Rutter, W.J. (1987). Selective alteration of substrate specificity by replacement of aspartic acid-189 with lysine in the binding pocket of trypsin. Biochemistry *26*, 2616-2623.
18. Hedstrom, L., Perona, J.J., and Rutter, W.J. (1994). Converting trypsin to chymotrypsin: residue 172 is a substrate specificity determinant. Biochemistry *33*, 8757.
19. Venekei, I., Szilágyi, L., Gráf, L., and Rutter, W.J. (1996). Attempts to convert chymotrypsin to trypsin. Febs Letters *379*, 143-147.
20. Wang, J.D., Herman, C., Tipton, K.A., Gross, C.A., and Weissman, J.S. (2002). Directed evolution of substrate-optimized GroEL/S chaperonins. Cell *111*, 1027-1039.
21. Sharabi, O., Peleg, Y., Mashiach, E., Vardy, E., Ashani, Y., Silman, I., Sussman, J.L., and Shifman, J.M. (2009). Design, expression and characterization of mutants of fasciculin optimized for interaction with its target, acetylcholinesterase. Protein Engineering Design and Selection.
22. Ricard, J., and Cornish-Bowden, A. (2005). Co-operative and allosteric enzymes: 20 years on. European Journal of Biochemistry *166*, 255-272.
23. Ambroggio, X.I., and Kuhlman, B. (2006). Design of protein conformational switches. Current opinion in structural biology *16*, 525-530.
24. Cui, Q., and Karplus, M. (2008). Allostery and cooperativity revisited. Protein Science: A Publication of the Protein Society *17*, 1295.
25. Dueber, J.E., Mirsky, E.A., and Lim, W.A. (2007). Engineering synthetic signaling proteins with ultrasensitive input/output control. Nature *200*, 7.
26. Fastrez, J. (2009). Engineering Allosteric Regulation into Biological Catalysts. Chembiochem *9999*.
27. May, L.T., Leach, K., Sexton, P.M., and Christopoulos, A. (2007). Allosteric modulation of G protein–coupled receptors.
28. Strickland, D., Moffat, K., and Sosnick, T.R. (2008). Light-activated DNA binding in a designed allosteric protein. Proceedings of the National Academy of Sciences *105*, 10709.
29. Yeh, B.J., Rutigliano, R.J., Deb, A., Bar-Sagi, D., and Lim, W.A. (2007). Rewiring cellular morphology pathways with synthetic guanine nucleotide exchange factors. Nature *447*, 596-600.
30. Popovych, N., Sun, S., Ebright, R.H., and Kalodimos, C.G. (2006). Dynamically driven protein allostery. Nature Structural & Molecular Biology *13*, 831.

31. Tsai, C.J., del Sol, A., and Nussinov, R. (2008). Allostery: absence of a change in shape does not imply that allostery is not at play. Journal of molecular biology *378*, 1-11.
32. Bindslev, N. (2008). Allostery and development of its models. Drug-Acceptor Interactions, 377.
33. Hilser, V.J. (2010). An Ensemble View of Allostery. Science *327*, 653-654.
34. H.M. Berman, J.W., Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne (2000). The Protein Data Bank. Nucleic Acids Res *28*, 235-242.
35. Ponder, J.W., and Richards, F.M. (1987). Tertiary templates for proteins. Use of packing criteria in the enumeration of allowed sequences for different structural classes. Journal of molecular biology *193*, 775.
36. Jaramillo, A., and Wodak, S.J. (2005). Computational protein design is a challenge for implicit solvation models. Biophysical journal *88*, 156-171.
37. Kiel, C., Selzer, T., Shaul, Y., Schreiber, G., and Herrmann, C. (2004). Electrostatically optimized Ras-binding Ral guanine dissociation stimulator mutants increase the rate of association by stabilizing the encounter complex. Proceedings of the National Academy of Sciences *101*, 9223.
38. Spiegel, K., De Grado, W.F., and Klein, M.L. (2006). Structural and dynamical properties of manganese catalase and the synthetic protein DF1 and their implication for reactivity from classical molecular dynamics calculations. Proteins: Structure, Function, and Bioinformatics *65*.
39. Saraf, M.C., Moore, G.L., Goodey, N.M., Cao, V.Y., Benkovic, S.J., and Maranas, C.D. (2006). IPRO: an iterative computational protein library redesign and optimization procedure. Biophysical journal *90*, 4167-4180.
40. Lippow, S.M., and Tidor, B. (2007). Progress in computational protein design. Current opinion in biotechnology *18*, 305-311.
41. Mandell, D.J., and Kortemme, T. (2009). Computer-aided design of functional protein interactions. Nature Chemical Biology *5*, 797-807.
42. Dahiyat, B.I., Sarisky, C.A., and Mayo, S.L. (1997). De novo protein design: towards fully automated sequence selection. Journal of molecular biology *273*, 789-796.
43. Getzoff, E.D., Cabelli, D.E., Fisher, C.L., Parge, H.E., Viezzoli, M.S., Banci, L., and Hallewell, R.A. (1992). Faster superoxide dismutase mutants designed by enhancing electrostatic guidance. Nature *358*, 347-351.
44. Cirino, P.C., Mayer, K.M., and Umeno, D. (2003). Generating mutant libraries using error-prone PCR. METHODS IN MOLECULAR BIOLOGY-CLIFTON THEN TOTOWA- *231*, 3-10.
45. Wong, T.S., Roccatano, D., Zacharias, M., and Schwaneberg, U. (2006). A statistical analysis of random mutagenesis methods used for directed protein evolution. Journal of molecular biology *355*, 858-871.
46. Patrick, W.M., Firth, A.E., and Blackburn, J.M. (2003). User-friendly algorithms for estimating completeness and diversity in randomized protein-encoding libraries. Protein Engineering Design and Selection *16*, 451.

47. Patrick, W.M., and Firth, A.E. (2005). Strategies and computational tools for improving randomized protein libraries. Biomolecular engineering *22*, 105-112.
48. Stemmer, W.P. (1994). DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. Proceedings of the National Academy of Sciences *91*, 10747.
49. Zhao, H., Giver, L., Shao, Z., Affholter, J.A., and Arnold, F.H. (1998). Molecular evolution by staggered extension process (StEP) in vitro recombination. Nature biotechnology *16*, 258-261.
50. Coco, W.M. (2003). RACHITT: Gene family shuffling by Random Chimeragenesis on Transient Templates. Methods in molecular biology (Clifton, NJ) *231*, 111.
51. Ostermeier, M., and Lutz, S. (2003). The creation of ITCHY hybrid protein libraries. METHODS IN MOLECULAR BIOLOGY-CLIFTON THEN TOTOWA- *231*, 129-142.
52. Joern, J.M., Meinhold, P., and Arnold, F.H. (2002). Analysis of shuffled gene libraries. Journal of molecular biology *316*, 643-656.
53. Neylon, C. (2004). Chemical and biochemical strategies for the randomization of protein encoding DNA sequences: library construction methods for directed evolution. Nucleic Acids Research *32*, 1448.
54. Daugherty, P.S. (2007). Protein engineering with bacterial display. Current opinion in structural biology *17*, 474-480.
55. Ellington, A.D., and Szostak, J.W. (1990). In vitro selection of RNA molecules that bind specific ligands. Nature *346*, 818-822.
56. Hall, B.G. (1981). Changes in the substrate specificities of an enzyme during directed evolution of new functions. Biochemistry *20*, 4042-4049.
57. Olsen, M., Iverson, B., and Georgiou, G. (2000). High-throughput screening of enzyme libraries. Current opinion in biotechnology *11*, 331-337.
58. Fields, S., and Song, O. (1989). A novel genetic system to detect protein protein interactions. Nature *340*, 245-246.
59. Vojtek, A.B., Hollenberg, S.M., and Cooper, J.A. (1993). Mammalian Ras interacts directly with the serine/threonine kinase Raf. Cell *74*, 205-214.
60. Durfee, T., Becherer, K., Chen, P.L., Yeh, S.H., Yang, Y., Kilburn, A.E., Lee, W.H., and Elledge, S.J. (1993). The retinoblastoma protein associates with the protein phosphatase type 1 catalytic subunit. Genes & development *7*, 555.
61. Gyuris, J., Golemis, E., Chertkov, H., and Brent, R. (1993). Cdi1, a human G1 and S phase protein phosphatase that associates with Cdk2. Cell *75*, 791-803.
62. Chien, C.T., Bartel, P.L., Sternglanz, R., and Fields, S. (1991). The two-hybrid system: a method to identify and clone genes for proteins that interact with a protein of interest. Proceedings of the National Academy of Sciences *88*, 9578.
63. Schneider, S., Buchert, M., Georgiev, O., Catimel, B., Halford, M., Stacker, S.A., Baechi, T., Moelling, K., and Hovens, C.M. (1999). Mutagenesis and selection of PDZ domains that bind new protein targets. Nature biotechnology *17*, 170-175.
64. Junqueira, D., Cilenti, L., Musumeci, L., Sedivy, J.M., and Zervos, A.S. (2003). Random mutagenesis of PDZOmi domain and selection of mutants that

specifically bind the Myc proto-oncogene and induce apoptosis. Oncogene *22*, 2772-2781.
65. Colas, P., and Brent, R. (1998). The impact of two-hybrid and related methods on biotechnology. Trends in Biotechnology *16*, 355-363.
66. Stellberger, T., Hauser, R., Baiker, A., Pothineni, V., Haas, J., and Uetz, P. (2010). Improving the yeast two-hybrid system with permutated fusions proteins: the Varicella Zoster Virus interactome. Proteome Science *8*, 8.
67. Smith, G.P. (1985). Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. Science *228*, 1315.
68. Lu, D., Shen, J., Vil, M.D., Zhang, H., Jimenez, X., Bohlen, P., Witte, L., and Zhu, Z. (2003). Tailoring in vitro selection for a picomolar-affinity human antibody directed against VEGF receptor 2 for enhanced neutralizing activity. J. Biol. Chem *12*, 12.
69. Hoogenboom, H.R. (2005). Selecting and screening recombinant antibody libraries. Nature biotechnology *23*, 1105-1116.
70. Ferrer, M., Maiolo, J., Kratz, P., Jackowski, J.L., Murphy, D.J., Delagrave, S., and Inglese, J. (2005). Directed evolution of PDZ variants to generate high-affinity detection reagents. Protein Engineering Design and Selection *18*, 165.
71. Yin, J., Mills, J.H., and Schultz, P.G. (2004). A catalysis-based selection for peroxidase antibodies with increased activity. Journal of the American Chemical Society *126*, 3006-3007.
72. Bradbury, A.R.M., and Marks, J.D. (2004). Antibodies from phage antibody libraries. Journal of immunological methods *290*, 29-49.
73. Sidhu, S.S., and Koide, S. (2007). Phage display for engineering and analyzing protein interaction interfaces. Current opinion in structural biology *17*, 481-487.
74. Tonikian, R., Zhang, Y., Boone, C., and Sidhu, S.S. (2007). Identifying specificity profiles for peptide recognition modules from phage-displayed peptide libraries. Nature Protocols *2*, 1368-1386.
75. Sidhu, S.S. (2000). Phage display in pharmaceutical biotechnology. Current opinion in biotechnology *11*, 610-616.
76. Ståhl, S., and Uhlén, M. (1997). Bacterial surface display: trends and progress. Trends in Biotechnology *15*, 185-192.
77. Li, M. (2000). Applications of display technology in protein analysis. Nature biotechnology *18*, 1251-1256.
78. Huh, D., Gu, W., Kamotani, Y., Grotberg, J.B., and Takayama, S. (2005). Review: Microfluidics for flow cytometric analysis of cells and particles. Physiological measurement *26*, R73-R98.
79. Agterberg, M., Adriaanse, H., and Tommassen, J. (1987). Use of outer membrane protein PhoE as a carrier for the transport of a foreign antigenic determinant to the cell surface of Escherichia coli K-12. Gene *59*, 145.
80. Charbit, A., Boulain, J.C., Ryter, A., and Hofnung, M. (1986). Probing the topology of a bacterial membrane protein by genetic insertion of a foreign epitope; expression at the cell surface. The EMBO Journal *5*, 3029.

81. Francisco, J.A., Campbell, R., Iverson, B.L., and Georgiou, G. (1993). Production and fluorescence-activated cell sorting of Escherichia coli expressing a functional antibody fragment on the external surface. Proc. Natl. Acad. Sci. USA *90*, 10444-10448.
82. Francisco, J.A., Earhart, C.F., and Georgiou, G. (1992). Transport and anchoring of beta-lactamase to the external surface of Escherichia coli. Proceedings of the National Academy of Sciences *89*, 2713.
83. Daugherty, P.S., Chen, G., Olsen, M.J., Iverson, B.L., and Georgiou, G. (1998). Antibody affinity maturation using bacterial surface display. Protein Engineering Design and Selection *11*, 825.
84. Daugherty, P.S., Chen, G., Iverson, B.L., and Georgiou, G. (2000). Quantitative analysis of the effect of the mutation frequency on the affinity maturation of single chain Fv antibodies. Proceedings of the National Academy of Sciences of the United States of America *97*, 2029.
85. Chen, G., Hayhurst, A., Thomas, J.G., Harvey, B.R., Iverson, B.L., and Georgiou, G. (2001). Isolation of high-affinity ligand-binding proteins by periplasmic expression with cytometric screening (PECS). Nature biotechnology *19*, 537-542.
86. Harvey, B.R., Georgiou, G., Hayhurst, A., Jeong, K.J., Iverson, B.L., and Rogers, G.K. (2004). Anchored periplasmic expression, a versatile technology for the isolation of high-affinity antibodies from Escherichia coli-expressed libraries. Proceedings of the National Academy of Sciences *101*, 9193.
87. Mabry, R., Rani, M., Geiger, R., Hubbard, G.B., Carrion Jr, R., Brasky, K., Patterson, J.L., Georgiou, G., and Iverson, B.L. (2005). Passive protection against anthrax by using a high-affinity antitoxin antibody fragment lacking an Fc region. Infection and immunity *73*, 8362.
88. Mazor, Y., Blarcom, T.V., Mabry, R., Iverson, B.L., and Georgiou, G. (2007). Isolation of engineered, full-length antibodies from libraries expressed in Escherichia coli. Nature biotechnology *25*, 563-565.
89. Seo, M.J., Jeong, K.J., Leysath, C.E., Ellington, A.D., Iverson, B.L., and Georgiou, G. (2009). Engineering antibody fragments to fold in the absence of disulfide bonds. Protein Science *18*, 259-267.
90. Boder, E.T., and Wittrup, K.D. (1997). Yeast surface display for screening combinatorial polypeptide libraries. Nature biotechnology *15*, 553-557.
91. Gai, S.A., and Wittrup, K.D. (2007). Yeast surface display for protein engineering and characterization. Current opinion in structural biology *17*, 467-473.
92. Chao, G., Lau, W.L., Hackel, B.J., Sazinsky, S.L., Lippow, S.M., and Wittrup, K.D. (2006). Isolating and engineering human antibodies using yeast surface display. Nat Protoc *1*, 755-768.
93. Scolnik, P.A. (2009). mAbs: A business perspective. mAbs *1*, 179-184.
94. Binz, H.K., Amstutz, P., and Pluckthun, A. (2005). Engineering novel binding proteins from nonimmunoglobulin domains. Nature biotechnology *23*, 1257-1268.
95. Skerra, A. (2007). Alternative non-antibody scaffolds for molecular recognition. Current opinion in biotechnology *18*, 295-304.

96. Kim, S.K. (1995). TIGHT JUNCTIONS, MEMBRANE-ASSOCIATED GUANYLATE KINASES AND CELL SIGNALING. Current Opinion in Cell Biology *7*, 641-649.
97. Cho, K.O., Hunt, C.A., and Kennedy, M.B. (1992). THE RAT-BRAIN POSTSYNAPTIC DENSITY FRACTION CONTAINS A HOMOLOG OF THE DROSOPHILA DISKS-LARGE TUMOR SUPPRESSOR PROTEIN. Neuron *9*, 929-942.
98. Woods, D.F., and Bryant, P.J. (1993). ZO-1, DLGA AND PSD-95/SAP90 - HOMOLOGOUS PROTEINS IN TIGHT, SEPTATE AND SYNAPTIC CELL-JUNCTIONS. Mechanisms of Development *44*, 85-89.
99. Kennedy, M.B. (1995). ORIGIN OF PDZ (DHR, GLGF) DOMAINS. Trends in Biochemical Sciences *20*, 350-350.
100. Ponting, C.P. (1997). Evidence for PDZ domains in bacteria, yeast, and plants. Protein Science *6*, 464-468.
101. Wiedemann, U., Boisguerin, P., Leben, R., Leitner, D., Krause, G., Moelling, K., Volkmer-Engert, R., and Oschkinat, H. (2004). Quantification of PDZ domain specificity, prediction of ligand affinity and rational design of super-binding peptides. Journal of molecular biology *343*, 703-718.
102. Songyang, Z., Fanning, A.S., Fu, C., Xu, J., Marfatia, S.M., Chishti, A.H., Crompton, A., Chan, A.C., Anderson, J.M., and Cantley, L.C. (1997). Recognition of unique carboxyl-terminal motifs by distinct PDZ domains. Science *275*, 73.
103. Saro, D., Martin, P., Vickrey, J.R., Griffin, A., Kovari, L.C., Spaller, M.R. (2009). Structure of the third PDZ domain of PSD-95 protein complexed with KKETPV peptide ligand from Rattus norvegicus. In Norris Cotton Cancer Center. (Lebanon, NH: Dartmouth College).
104. DeLano, W.L. (2002). The PyMOL User's Manual, (Palo Alto, CA).
105. Fan, J.S., and Zhang, M. (2000). Signaling complex organization by PDZ domain proteins. Neurosignals *11*, 315-321.
106. Harris, B.Z., Hillier, B.J., and Lim, W.A. (2001). Energetic determinants of internal motif recognition by PDZ domains. Biochemistry *40*, 5921-5930.
107. Zimmermann, P. (2006). The prevalence and significance of PDZ domain–phosphoinositide interactions. BBA-Molecular and Cell Biology of Lipids *1761*, 947-956.
108. Nourry, C., Grant, S.G.N., and Borg, J.P. (2003). PDZ domain proteins: plug and play! Science Signaling *2003*.
109. Ludford-Menting, M.J., Oliaro, J., Sacirbegovic, F., Cheah, E.T.Y., Pedersen, N., Thomas, S.J., Pasam, A., Iazzolino, R., Dow, L.E., and Waterhouse, N.J. (2005). A network of PDZ-containing proteins regulates T cell polarity and morphology during migration and immunological synapse formation. Immunity *22*, 737-748.
110. Brone, B., and Eggermont, J. (2005). PDZ proteins retain and regulate membrane transporters in polarized epithelial cell membranes. American Journal of Physiology-Cell Physiology *288*, C20-C29.

111. Fanning, A.S., and Anderson, J.M. (1996). Protein-protein interactions: PDZ domain networks. Current Biology *6*, 1385-1388.
112. Fanning, A.S., and Anderson, J.M. (1999). PDZ domains: fundamental building blocks in the organization of protein complexes at the plasma membrane. Journal of Clinical Investigation *103*, 767-772.
113. Garner, C.C., Nash, J., and Huganir, R.L. (2000). PDZ domains in synapse assembly and signalling. Trends in Cell Biology *10*, 274-280.
114. Hung, A.Y., and Sheng, M. (2002). PDZ domains: Structural modules for protein complex assembly. Journal of Biological Chemistry *277*, 5699-5702.
115. Jankun, J., Doerks, T., Aleem, A.M., Lysiak-Szydlowska, W., and Skrzypczak-Jankun, E. (2008). Do Human Lipoxygenases have a PDZ Regulatory Domain? Current Molecular Medicine *8*, 768-773.
116. Jelen, F., Oleksy, A., Smietana, K., and Otlewski, J. (2003). PDZ domains - common players in the cell signaling. Acta Biochimica Polonica *50*, 985-1017.
117. Kim, E.J., and Sheng, M. (2004). PDZ domain proteins of synapses. Nature Reviews Neuroscience *5*, 771-781.
118. Tsunoda, S., Sierralta, J., Sun, Y., Bodner, R., Suzuki, E., Becker, A., Socolich, M., and Zuker, C.S. (1997). A multivalent PDZ-domain protein assembles signalling complexes in a G-protein-coupled cascade. Nature *388*, 243-243.
119. Aarts, M., Liu, Y., Liu, L., Besshoh, S., Arundine, M., Gurd, J.W., Wang, Y.T., Salter, M.W., and Tymianski, M. (2002). Treatment of ischemic brain damage by perturbing NMDA receptor-PSD-95 protein interactions. Science Signaling *298*, 846.
120. Alto, N.M., Shao, F., Lazar, C.S., Brost, R.L., Chua, G., Mattoo, S., McMahon, S.A., Ghosh, P., Hughes, T.R., and Boone, C. (2006). Identification of a bacterial type III effector family with G protein mimicry functions. Cell *124*, 133-145.
121. Beresford, I.J.M., Parsons, A.A., and Hunter, A.J. (2003). Treatments for stroke. emd *8*, 103-122.
122. Dev, K.K. (2004). Making protein interactions druggable: targeting PDZ domains. Nature Reviews Drug Discovery *3*, 1047-1056.
123. Doorbar, J. (2006). Molecular biology of human papillomavirus infection and cervical cancer. Clinical Science *110*, 525-541.
124. Frese, K.K., Lee, S.S., Thomas, D.L., Latorre, I.J., Weiss, R.S., Glaunsinger, B.A., and Javier, R.T. (2003). Selective PDZ protein-dependent stimulation of phosphatidylinositol 3-kinase by the adenovirus E4-ORF1 oncoprotein. Oncogene *22*, 710-721.
125. Hall, W.W., and Fujii, M. (2005). Deregulation of cell-signaling pathways in HTLV-1 infection. Oncogene *24*, 5965-5975.
126. Humbert, P., Russell, S., and Richardson, H. (2003). Dlg, Scribble and Lgl in cell polarity, cell proliferation and cancer. Bioessays *25*, 542-553.
127. Obenauer, J.C., Denson, J., Mehta, P.K., Su, X., Mukatira, S., Finkelstein, D.B., Xu, X., Wang, J., Ma, J., and Fan, Y. (2006). Large-scale sequence analysis of avian influenza isolates. Science *311*, 1576.

128. Uematsu, K., Kanazawa, S., You, L., He, B., Xu, Z., Li, K., Peterlin, B.M., McCormick, F., and Jablons, D.M. (2003). Wnt Pathway Activation in Mesothelioma: Evidence of Dishevelled Overexpression and Transcriptional Activity of {beta}-Catenin. Cancer research *63*, 4547.
129. Aasland, R., Abrams, C., Ampe, C., Ball, L.J., Bedford, M.T., Cesareni, G., Gimona, M., Hurley, J.H., Jarchau, T., and Lehto, V.P. (2002). Normalization of nomenclature for peptide motifs as ligands of modular protein domains. Febs Letters *513*, 141-144.
130. Stricker, N.L., Christopherson, K.S., Yi, B.A., Schatz, P.J., Raab, R.W., Dawes, G., Bassett, D.E., Bredt, D.S., and Li, M. (1997). PDZ domain of neuronal nitric oxide synthase recognizes novel C-terminal peptide sequences. Nature biotechnology *15*, 336-342.
131. Bezprozvanny, I., and Maximov, A. (2001). Classification of PDZ domains. Febs Letters *509*, 457-462.
132. Bezprozvanny, I., and Maximov, A. (2002). PDZ domains: evolving classification. Febs Letters *512*, 347-349.
133. Vaccaro, P., and Dente, L. (2002). PDZ domains: troubles in classification. Febs Letters *512*, 345-346.
134. Kang, B.S., Cooper, D.R., Devedjiev, Y., Derewenda, U., and Derewenda, Z.S. (2003). Molecular Roots of Degenerate Specificity in Syntenin's PDZ2 Domain Reassessment of the PDZ Recognition Paradigm. Structure *11*, 845-853.
135. Chen, J.R., Chang, B.H., Allen, J.E., Stiffler, M.A., and MacBeath, G. (2008). Predicting PDZ domain–peptide interactions from primary sequences. Nature biotechnology *26*, 1041.
136. Tonikian, R., Zhang, Y., Sazinsky, S.L., Currell, B., Yeh, J.H., Reva, B., Held, H.A., Appleton, B.A., Evangelista, M., and Wu, Y. (2008). A specificity map for the PDZ domain family. PLoS Biol *6*, e239.
137. Reina, J., Lacroix, E., Hobson, S.D., Fernandez-Ballester, G., Rybin, V., Schwab, M.S., Serrano, L., and Gonzalez, C. (2002). Computer-aided design of a PDZ domain to recognize new target sequences. nature structural biology *9*, 621-627.
138. Boettner, B., Govek, E.E., Cross, J., and Van Aelst, L. (2000). The junctional multidomain protein AF-6 is a binding partner of the Rap1A GTPase and associates with the actin cytoskeletal regulator profilin. Proceedings of the National Academy of Sciences of the United States of America *97*, 9064.
139. Yamamoto, T., Harada, N., Kano, K., Taya, S., Canaani, E., Matsuura, Y., Mizoguchi, A., Ide, C., and Kaibuchi, K. (1997). The Ras target AF-6 interacts with ZO-1 and serves as a peripheral component of tight junctions in epithelial cells. Journal of Cell Biology *139*, 785.
140. Fisinger, S., Serrano, L., and Lacroix, E. (2001). Computational estimation of specific side chain interaction energies in helices. Protein Science: A Publication of the Protein Society *10*, 809.
141. Lacroix, E., Ramirez-Alvarado, M., and Serrano, L. (2001). Computer-aided Design of beta-sheet Peptides. Journal of molecular biology *312*, 229-246.

142. Amati, B., Littlewood, T.D., Evan, G.I., and Land, H. (1993). The c-Myc protein induces cell cycle progression and apoptosis through dimerization with Max. The EMBO Journal *12*, 5083.
143. Hamilton, A.C., Inglese, J., and Ferrer, M. (2003). A PDZ domain-based assay for measuring HIV protease activity: Assay design considerations. Protein Science: A Publication of the Protein Society *12*, 458.
144. Swain, J.F., and Gierasch, L.M. (2006). The changing landscape of protein allostery. Current opinion in structural biology *16*, 102-108.
145. Gunasekaran, K., Ma, B., and Nussinov, R. (2004). Is allostery an intrinsic property of all dynamic proteins. Proteins: Struct. Funct. Bioinf *57*, 433–443.
146. Smock, R.G., and Gierasch, L.M. (2009). Sending signals dynamically. Science Signaling *324*, 198.
147. Cori, G.T., Colowick, S.P., and Cori, C.F. (1938). The action of nucleotides in the disruptive phosphorylation of glycogen. Journal of Biological Chemistry *123*, 381-389.
148. Lindsley, J.E., and Rutter, J. (2006). Whence cometh the allosterome? Proceedings of the National Academy of Sciences *103*, 10533.
149. Monod, J., Changeux, J.P., and Jacob, F. (1963). Allosteric proteins and cellular control systems. Journal of molecular biology *6*, 306.
150. Monod, J., Wyman, J., and Changeux, J.P. (1965). On the nature of allosteric transitions: a plausible model. Journal of molecular biology *12*, 88.
151. Koshland Jr, D.E., N methy, G., and Filmer, D. (1966). Comparison of Experimental Binding Data and Theoretical Models in Proteins Containing Subunits*. Biochemistry *5*, 365-385.
152. Suel, G.M., Lockless, S.W., Wall, M.A., and Ranganathan, R. (2003). Evolutionarily conserved networks of residues mediate allosteric communication in proteins. nature structural biology *10*, 59-69.
153. Fuentes, E.J., Der, C.J., and Lee, A.L. (2004). Ligand-dependent dynamics and intramolecular signaling in a PDZ domain. Journal of molecular biology *335*, 1105-1115.
154. Ota, N., and Agard, D.A. (2005). Intramolecular signaling pathways revealed by modeling anisotropic thermal diffusion. Journal of molecular biology *351*, 345-354.
155. Koide, S. (2009). Generation of new protein functions by nonhomologous combinations and rearrangements of domains and modules. Current opinion in biotechnology *20*, 398-404.
156. Ostermeier, M. (2009). Designing switchable enzymes. Current opinion in structural biology *19*, 442-448.
157. Wright, C.M., Heins, R.A., and Ostermeier, M. (2007). As easy as flipping a switch? Current Opinion in Chemical Biology *11*, 342-346.
158. Bashor, C.J., Helman, N.C., Yan, S., and Lim, W.A. (2008). Using engineered scaffold interactions to reshape MAP kinase pathway signaling dynamics. Science *319*, 1539.

159. Dueber, J.E., Yeh, B.J., Chak, K., and Lim, W.A. (2003). Reprogramming control of an allosteric signaling switch through modular recombination. Science *301*, 1904.
160. Pérez-Rueda, E., and Collado-Vides, J. (2000). The repertoire of DNA-binding transcriptional regulators in Escherichia coli K-12. Nucleic Acids Research *28*, 1838.
161. Gruber, T.M., and Gross, C.A. (2003). M ULTIPLE S IGMA S UBUNITS AND THE P ARTITIONING OF B ACTERIAL T RANSCRIPTION S PACE. Annual Reviews in Microbiology *57*, 441-466.
162. Erickson, J.W., and Gross, C.A. (1989). Identification of the sigma E subunit of Escherichia coli RNA polymerase: a second alternate sigma factor involved in high-temperature gene expression. Genes & development *3*, 1462.
163. Hayden, J.D., and Ades, S.E. (2008). The Extracytoplasmic Stress Factor, E, Is Required to Maintain Cell Envelope Integrity in Escherichia coli. PLoS One *3*.
164. Thompson, K.M., Rhodius, V.A., and Gottesman, S. (2007). {sigma} E regulates and is regulated by a small RNA in Escherichia coli. Journal of bacteriology *189*, 4243.
165. Costanzo, A., Nicoloff, H., Barchinger, S., Banta, A., Gourse, R., and Ades, S. (2008). ppGpp and DksA likely regulate the activity of the extracytoplasmic stress factor sE in Escherichia coli by both direct and indirect mechanisms. Molecular Microbiology *67*, 619-632.
166. Alba, B.M., Zhong, H.J., Pelayo, J.C., and Gross, C.A. (2001). degS (hhoB) is an essential Escherichia coli gene whose indispensable function is to provide sigma^ E activity. Molecular Microbiology *40*, 1323-1333.
167. Ades, S.E. (2004). Control of the alternative sigma factor E in Escherichia coli. Current opinion in microbiology *7*, 157-162.
168. Grigorova, I.L., Chaba, R., Zhong, H.J., Alba, B.M., Rhodius, V., Herman, C., and Gross, C.A. (2004). Fine-tuning of the Escherichia coli E envelope stress response relies on multiple mechanisms to inhibit signal-independent proteolysis of the transmembrane anti-sigma factor, RseA. Genes & development *18*, 2686.
169. Costanzo, A., and Ades, S.E. (2006). Growth phase-dependent regulation of the extracytoplasmic stress factor,{sigma} E, by guanosine 3', 5'-bispyrophosphate (ppGpp). Journal of bacteriology *188*, 4627.
170. Chaba, R., Grigorova, I.L., Flynn, J.M., Baker, T.A., and Gross, C.A. (2007). Design principles of the proteolytic cascade governing the E-mediated envelope stress response in Escherichia coli: keys to graded, buffered, and rapid signal transduction. Genes & development *21*, 124.
171. Ades, S.E., Connolly, L.E., Alba, B.M., and Gross, C.A. (1999). The Escherichia coli E-dependent extracytoplasmic stress response is controlled by the regulated proteolysis of an anti- factor. Genes & development *13*, 2449.
172. Walsh, N.P., Alba, B.M., Bose, B., Gross, C.A., and Sauer, R.T. (2003). OMP peptide signals initiate the envelope-stress response by activating DegS protease via relief of inhibition mediated by its PDZ domain. Cell *113*, 61-71.

173. Alba, B.M., Leeds, J.A., Onufryk, C., Lu, C.Z., and Gross, C.A. (2002). DegS and YaeL participate sequentially in the cleavage of RseA to activate the E-dependent extracytoplasmic stress response. Genes & development *16*, 2156.
174. Ruiz, N., and Silhavy, T.J. (2005). Sensing external stress: watchdogs of the Escherichia coli cell envelope. Current opinion in microbiology *8*, 122-126.
175. Ades, S.E. (2008). Regulation by destruction: design of the sigma(E) envelope stress response. Current opinion in microbiology *11*, 535-540.
176. Heimann, J.D. (2002). The extracytoplasmic function(ECF) sigma factors. Advances in microbial physiology *46*, 47-110.
177. Raivio, T.L., and Silhavy, T.J. (2001). P ERIPLASMIC S TRESS AND ECF S IGMA F ACTORS. Annual Reviews in Microbiology *55*, 591-624.
178. Ades, S.E. (2008). Regulation by destruction: design of the E envelope stress response. Current opinion in microbiology *11*, 535-540.
179. Ades, S.E., Grigorova, I.L., and Gross, C.A. (2003). Regulation of the alternative sigma factor {sigma} E during initiation, adaptation, and shutoff of the extracytoplasmic heat shock response in Escherichia coli. Journal of bacteriology *185*, 2512.
180. Alba, B.M., and Gross, C.A. (2004). Regulation of the Escherichia colisigmaE-dependent envelope stress response. Molecular Microbiology *52*, 613-619.
181. Hasselblatt, H., Kurzbauer, R., Wilken, C., Krojer, T., Sawa, J., Kurt, J., Kirk, R., Hasenbein, S., Ehrmann, M., and Clausen, T. (2007). Regulation of the E stress response by DegS: how the PDZ domain keeps the protease inactive in the resting state and allows integration of different OMP-derived stress signals upon folding stress. Genes & development *21*, 2659.
182. Sohn, J., Grant, R.A., and Sauer, R.T. (2007). Allosteric activation of DegS, a stress sensor PDZ protease. Cell *131*, 572-583.
183. Sohn, J., and Sauer, R.T. (2009). OMP peptides modulate the activity of DegS protease by differential binding to active and inactive conformations. Molecular Cell *33*, 64-74.
184. Wilken, C., Kitzing, K., Kurzbauer, R., Ehrmann, M., and Clausen, T. (2004). Crystal Structure of the DegS Stress Sensor How a PDZ Domain Recognizes Misfolded Protein and Activates a Protease. Cell *117*, 483-494.
185. Hauske, P., Mamant, N., Hasenbein, S., Nickel, S., Ottmann, C., Clausen, T., Ehrmann, M., and Kaiser, M. (2009). Peptidic small molecule activators of the stress sensor DegS. Molecular BioSystems *5*, 980-985.
186. Sohn, J., Grant, R.A., and Sauer, R.T. (2009). OMP Peptides Activate the DegS Stress-Sensor Protease by a Relief of Inhibition Mechanism. Structure *17*, 1411-1421.
187. Zeth, K. (2004). Structural analysis of DegS, a stress sensor of the bacterial periplasm. Febs Letters *569*, 351-358.
188. Nooren, I.M.A., and Thornton, J.M. (2003). New embo member's review: diversity of protein–protein interactions. The EMBO Journal *22*, 3486.

189. Saro, D., Klosi, E., Paredes, A., and Spaller, M.R. (2004). Thermodynamic analysis of a hydrophobic binding site: probing the PDZ domain with nonproteinogenic peptide ligands. Org. Lett *6*, 3429-3432.
190. Fromant, M., Blanquet, S., and Plateau, P. (1995). Direct random mutagenesis of gene-sized DNA fragments using polymerase chain reaction. Analytical biochemistry *224*, 347-353.
191. Sarkar, C.A., Dodevski, I., Kenig, M., Dudli, S., Mohr, A., Hermans, E., and Plückthun, A. (2008). Directed evolution of a G protein-coupled receptor for expression, stability, and binding selectivity. Proceedings of the National Academy of Sciences *105*, 14808.
192. Varadarajan, N., Gam, J., Olsen, M.J., Georgiou, G., and Iverson, B.L. (2005). Engineering of protease variants exhibiting high catalytic activity and exquisite substrate selectivity. Proceedings of the National Academy of Sciences *102*, 6855.
193. Lee, C., and Laimins, L.A. (2004). Role of the PDZ domain-binding motif of the oncoprotein E6 in the pathogenesis of human papillomavirus type 31. The Journal of Virology *78*, 12366.
194. Hale, B.G., Randall, R.E., Ortin, J., and Jackson, D. (2008). The multifunctional NS1 protein of influenza A viruses. Journal of General Virology *89*, 2359.
195. Ostermeier, M. (2005). Engineering allosteric protein switches by domain insertion. Protein Engineering Design and Selection *18*, 359.
196. Guntas, G., Mansell, T.J., Kim, J.R., and Ostermeier, M. (2005). Directed evolution of protein switches and their application to the creation of ligand-binding proteins. Proceedings of the National Academy of Sciences of the United States of America *102*, 11224.
197. Koonin, E.V., and Aravind, L. (2002). Origin and evolution of eukaryotic apoptosis: the bacterial connection. Cell death and differentiation *9*, 394-404.
198. Heinrich, J., and Wiegert, T. (2009). Regulated intramembrane proteolysis in the control of extracytoplasmic function sigma factors. Research in Microbiology *160*, 696-703.
199. Challa, M., Malladi, S., Pellock, B.J., Dresnek, D., Varadarajan, S., Yin, Y.W., White, K., and Bratton, S.B. (2007). Drosophila Omi, a mitochondrial-localized IAP antagonist and proapoptotic serine protease. The EMBO Journal *26*, 3144.
200. Clausen, T., Southan, C., and Ehrmann, M. (2002). The HtrA Family of Proteases Implications for Protein Composition and Cell Fate. Molecular Cell *10*, 443-455.
201. Higgins, D., Thompson, J., Gibson, T., Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res *22*, 4673-4680.
202. Waller, P.R., and Sauer, R.T. (1996). Characterization of degQ and degS, Escherichia coli genes encoding homologs of the DegP protease. Journal of bacteriology *178*, 1146.
203. Meltzer, M., Hasenbein, S., Mamant, N., Merdanovic, M., Poepsel, S., Hauske, P., Kaiser, M., Huber, R., Krojer, T., and Clausen, T. (2009). Structure, function

and regulation of the conserved serine proteases DegP and DegS of Escherichia coli. Research in Microbiology.
204. Hauske, P., Meltzer, M., Ottmann, C., Krojer, T., Clausen, T., Ehrmann, M., and Kaiser, M. (2009). Selectivity profiling of DegP substrates and inhibitors. Bioorganic & medicinal chemistry *17*, 2920-2924.
205. Krojer, T., Garrido-Franco, M., Huber, R., Ehrmann, M., and Clausen, T. (2002). Crystal structure of DegP (HtrA) reveals a new protease-chaperone machine. Nature *416*, 455-459.
206. Krojer, T., Sawa, J., Schäfer, E., Saibil, H.R., Ehrmann, M., and Clausen, T. (2008). Structural basis for the regulated protease and chaperone function of DegP. Nature *453*, 885-890.
207. Kolmar, H., Waller, P.R., and Sauer, R.T. (1996). The DegP and DegQ periplasmic endoproteases of Escherichia coli: specificity for cleavage sites and substrate conformation. Journal of bacteriology *178*, 5925.
208. Kim, D.Y., and Kim, K.K. (2005). Structure and function of HtrA family proteins, the key players in protein quality control. Journal of Biochemistry and Molecular Biology *38*, 266-274.
209. Walle, L.V., Lamkanfi, M., and Vandenabeele, P. (2008). The mitochondrial serine protease HtrA2/Omi: an overview. Cell Death & Differentiation *15*, 453-460.
210. Kooistra, J., Milojevic, J., Melacini, G., and Ortega, J. (2009). A New Function of Human HtrA2 as an Amyloid- Oligomerization Inhibitor. Journal of Alzheimer's Disease *17*, 281-294.
211. Runyon, S.T., Zhang, Y., Appleton, B.A., Sazinsky, S.L., Wu, P., Pan, B., Wiesmann, C., Skelton, N.J., and Sidhu, S.S. (2007). Structural and functional analysis of the PDZ domains of human HtrA1 and HtrA3. Protein Science: A Publication of the Protein Society *16*, 2454.

VITA

Marsha Jane Lewis was born in Pontiac, Michigan, the daughter of Louise and the late Edward Demers and the oldest of five children. After attending public schools in New Jersey and Illinois and joining the Illinois Army National Guard as a combat medic, she matriculated into the University of Illinois at Urbana-Champaign. While at the University of Illinois, she entered a cooperative work agreement with Dow Chemical Co. and conducted undergraduate research through the honors program with Professor Thomas J. Hanratty investigating heat transfer across turbulent flow between two plates for materials characteristic of three Prandtl numbers. She received a Bachelor of Sciences degree with distinction in Chemical Engineering in December 1993.

In 1994, she was employed by Merck and Co, Inc. as a technical services engineer and was promoted consistently; successfully executing multiple roles and culminating as a Factory Department Manager that lead a manufacturing team, which produced bulk active pharmaceutical ingredients for the last two years of employment in manufacturing. After Merck and Co, Inc., she entered the University of Texas at Austin graduate program in the Institute of Cellular and Molecular Biology to pursue her doctorate degree. While in the graduate program, she married and had one daughter.

Permanent Address: 1559 Red Town Rd., Elgin, TX 78921.

This manuscript was typed by the author.