2013

Investigating the Interaction of Zinc-Finger Peptides and Nicked DNA Using Phage Display

Rachel Maria Guerra
Colby College

Follow this and additional works at: https://digitalcommons.colby.edu/honorstheses

Part of the Chemistry Commons

Recommended Citation

Investigating the Interaction of Zinc-Finger Peptides and Nicked DNA using Phage Display

By Rachel Maria Guerra

A Thesis Presented to the Department of Chemistry,
Colby College, Waterville, ME
In Partial Fulfillment of the Requirements for Graduation
With Honors in Chemistry

Submitted May, 2013

Vitae

Rachel Maria Guerra was born in Milton, Massachusetts to Rose and Richard Guerra. She completed her high school education at Ursuline Academy in Dedham, Massachusetts and graduated in 2009. Rachel attended Colby College in Waterville, Maine for her undergraduate studies, and she will graduate on May 26th, 2013 with a Bachelors of Arts degree in Chemistry-Biochemistry. Rachel will be pursuing graduate studies at Harvard University to earn her Doctor of Philosophy in Biological and Biomedical Sciences.

Permanent Address
45 Park Street
Milton, MA 02186

Acknowledgements

I would like to thank Professor Kevin Rice for providing me with the opportunity to work in his lab. I am so grateful for his constant mentoring and guidance over the past four years. He not only was incredibly helpful in expanding my research horizons, from summer internships to graduate school, but was also instrumental to my growth as a scientist. I would also like to thank Professor Julie Millard for reading my thesis and helping me develop my critical thinking skills in her biochemistry course. I am also very grateful for all of the help and support that Tara Kraus has provided me over the years, especially with troubleshooting. Thank you also to the members of the Rice Lab, both past and present, who have truly made me look forward to coming into lab even during the times when my project was most challenging and frustrating. Finally, I would like to thank my family and friends for their unfailing love and encouragement during the past four years. This work was supported by grants from the Colby College Division of Natural Sciences.

Table of Contents

Abstract

Recognition of structural features of DNA such as gaps, nicks and abasic sites is critical for many DNA-binding proteins, such as those involved in DNA damage repair. This study explores a novel strategy to model these protein-DNA interactions using phage display. Phage display has been commonly used to identity zinc fingers that bind to DNA sequences, but selection of peptides that binding to specific DNA structures has not been reported in the literature. A phage library of 10^9 variants was created based on one of three zinc finger domains of the DNA repair protein poly(ADP-ribose) polymerase I. The library, engineered at the DNA level using synthetic oligonucleotides and containing eight randomized codons in the "loop" region of the zinc finger, was cloned into a commercial T7 phagemid vector. Four rounds of selection against an immobilized nicked DNA target were carried out, the last of which resulted in 62% retention of phage. Sequencing of selected phage from this highly enriched library revealed convergence to a single 39-mer peptide, a truncated version of the designed 85-mer zinc finger. This peptide will be purified from transformed bacterial culture and its binding activity will be measured using fluorescence polarization. This project seeks to provide insight into the structure-function relationship of proteins that bind specific DNA structures and could possibly lead to a small peptide with diagnostic or therapeutic applications.

Introduction

Proteins are principally responsible for carrying out the metabolic and biochemical pathways necessary to sustain life. In addition to their enzymatic roles, proteins mediate macromolecular recognition events that trigger critical biological phenomena. Among these are the protein-protein interactions, as well as those between proteins and DNA. Protein-protein recognition is incredibly diverse, ranging from long-lived, rigid-body binding of antigen-antibody complexes, to more flexible interactions of proteases and inhibitors [1]. Regardless of the biological nature of the interaction, proteins can have exquisite specificity at the macromolecular level as a result of highly conserved residues at the sequence level, as well as specific structural and spatial physical and chemical properties [2]. Given this great diversity, protein-protein and protein-DNA interactions drive critical pathways in cells in a concerted manner. Transcription alone requires the macromolecular interactions of over 40 proteins [3]. Macromolecular interfaces are also of great interest in the drug discovery, as an increasing number of pharmaceuticals are designed to inhibit protein-protein and protein-DNA interactions [4]. Studying these macromolecular interactions is not only important for basic understanding of cellular processes at a molecular level, but is also key for allowing researchers to manipulate these interactions for practical applications such as drug discovery.

Protein-nucleic acid recognition is also incredibly important to biological processes, particularly gene expression, transcription, and translation. The ability of transcription factors to regulate gene expression stems from their recognition of specific DNA sequences; however, they also are characterized by flexibility and redundancy, which allows them to recognize similar sequences throughout the genome [5]. Hydrogen bonding between amino acids and DNA bases allows for recognition and discrimination of different bases; however, there is no universal set of

rules that dictates base-amino acid pairing [5, 6]. The complementarity of DNA binding sites in proteins reflects the distinctive chemical features of DNA, namely the highly anionic phosphate backbone [1]. Other important aspects to protein-nucleic acid recognition include the conformational flexibility of the binding interactions and cooperativity between multiple proteins and DNA or RNA [5]. In addition to the recognition of DNA sequences needed for regulation of gene expression, it is crucial that proteins are also able to recognize structural features of DNA, such as intermediates of DNA damage. The DNA within each cell is constantly subjected to DNA damaging events from a variety of sources. These events can occur during normal cell metabolic activities such as replication, in which DNA is often nicked, as well as a result of exposure reactive oxygen species, UV light, and ionizing radiation [7]. It is estimated that human cells experience approximately 100,000 single-strand DNA breaks per day, in addition to numerous incidents of base loss, alkylation, oxidation, and dimerization [8]. Repair of DNA damage is crucial to securing the integrity of the genome, and failure to repair the damage can result in disease, most notably cancer. Acquired mutations in tumor suppressor genes and proto-oncogenes can result in the transformation of normal cells into malignant cancer cells [9]. Human cells are equipped with numerous proteins and mechanisms responsible for recognizing structural features of damaged DNA and repairing that damage [10]. Macromolecular interactions of proteins with DNA not only are critical in gene expression, but also are key to protecting cells from both exogenous and endogenous DNA damage events.

DNA-binding proteins contain designated domains that recognize and bind DNA: the three most common categories are helix-turn-helix, zinc- leucine-zipper, and zinc-binding domains. These proteins are quite abundant, as genes for DNA-binding proteins comprise 6-7% of the human genome [11]. Helix-turn-helix (HTH) domains generally contain 20 residues, which coil into two α-helices, connected by a 4-residue β turn, that intersect at a 120° angle [12].

The second helix, or recognition helix, of the HTH binds to the major groove of DNA and is critical for HTH specificity [13, 14]. The second major class of DNA-domains is the leucine-zipper. The leucine-zipper contains two main regions: a 30-residue C-terminal dimerization region, and a 30-residue basic region at the N-terminal [11]. The dimerization region, or zipper region, contains regularly repeating leucine residues which allow for the formation of an α-helical coiled-coil during dimerization [12]. In the absence of DNA, the basic region is unstructured; however, it coils upon binding to the DNA major groove [14]. Proteins with zinc-binding domains are the most prominent class of DNA-binding proteins in the human genome, primarily as transcription factors [11]. There are multiple types of zinc-binding domains; however, coordination of Zn^{2+} is necessary for structural stability in all classes [15]. The prototypical zinc finger is approximately 30 residues, with one zinc ion coordinated by two conserved cysteines and two histidines; however, very few other residues are fully conserved [12, 15]. Other zinc-binding domains have two zinc ions coordinated by four cysteines or by six cysteines [12]. Zinc fingers generally bind a specific three base pair site and insert a helical region in the DNA major groove [12, 14]. Proteins with multiple zinc finger domains will wrap around the DNA and form a spiral [11]. Other types of DNA-binding domains exist outside of these major classes, including helix-loop-helix, β-sheets, and loop structures [12, 14]. Despite the diversity of DNA-binding domains, all classes have commonalities in their specificity of DNA binding, which is derived from a combination of hydrogen bonding, van der Waals interactions, and solvation effects [14]. By stringing together a series of DNA-binding domains, proteins are able to achieve exquisite specificity in recognizing DNA sequences.

Similar DNA-binding domains can also recognize structural features of DNA. Domains that bind structural features of DNA are often found in DNA repair enzymes, including those associated with base excision repair, nucleotide excision repair, mismatch repair, and double-

strand break repair [16]. DNA lesions of just one base pair, including nicks, gaps, and mismatched bases, often do not significantly disrupt the double-helix structure; thus, protein recognition relies on destabilized base stacking forces [16]. In particular, zinc fingers in the poly(ADP-ribose) polymerase-like zinc finger family are able to recognize similar structural features of DNA [17]. These zinc fingers are found in DNA repair enzymes, including PARPs, mammalian DNA ligase III, and plant DNA 3'-phosphatases, and are able to recognize varied secondary structures of DNA, such as DNA nicks, gaps, hairpins, double-strand breaks, and cruciforms [17]. The binding of the PARP-like zinc finger in DNA ligase III to DNA gaps, flaps, and nicks stimulates the end-joining activity of Lig III in DNA repair [18, 19]. These domains that bind to structural features of DNA are critical to the enzymatic activity of many DNA repair enzymes, and the study of these interactions is of importance due to the relationship between DNA damage repair and oncogenesis.

The zinc finger domains of PARP-1 provide an interesting glimpse into the macromolecular interaction of proteins and structural features of DNA. PARP-1 is one of eighteen currently known members of the PARP enzyme family, which serves important roles in the cell, most importantly repair of single-strand DNA breaks through base excision repair [20]. Upon recognizing damaged DNA, PARP-1 uses NAD^+ to ADP-ribosylate histones, nuclear proteins, and itself in order to sequester other components involved in DNA repair [7]. These base excision repair proteins include XRCC1, DNA polymerase β, APE-1, and DNA ligase III [21]. It is now been experimentally shown that PARP-1 recognizes a variety of DNA structural features; however, it has often been termed as a "molecular nick sensor" due to its ability to located nicked DNA [20]. PARP-1 contains three zinc fingers, ZnF1, ZnF2, and ZnF3, which are responsible for binding to DNA and activating enzymatic activity [22].
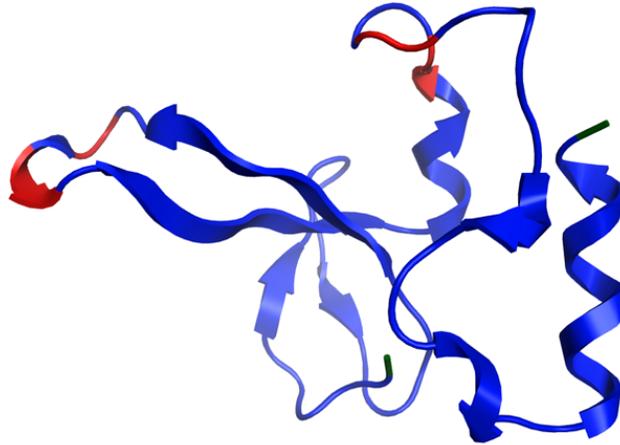
Figure 1. A representation of the ZnF2 domain of PARP (PDB: 4AV1).

In addition to the two conserved cysteine and histidine residues found in all zinc fingers, the PARP zinc finger domain contains four conserved hydrophobic and four conserved charged amino acids, which are believed to be essential for DNA binding [17]. Variable regions, however, may be responsible for imparting binding affinity with different DNA structural specificity to different groups of PARP-like zinc fingers [23]. As seen in Figure 1, the zinc fingers also contain two α-helical regions that are believed to be responsible for binding to other zinc fingers, as well as a triple-stranded β-sheet and two loop regions that form the binding interface with damaged DNA [24].

Current models of PARP-1 binding to DNA propose that the two zinc fingers, ZnF1 and ZnF2, bind cooperatively in opposite grooves of damaged DNA [25]. The three PARP zinc fingers, however, are thought to each have a unique function. ZnF2 is required for binding affinity of damaged DNA, while ZnF1 is necessary for the DNA-dependent activation of the enzymatic activity of PARP-1 [22]. Additionally, the zinc fingers display slightly different binding interactions with DNA, as ZnF1 makes contact primarily through the base stacking loop, whereas the binding interface of ZnF2 with DNA is a phosphate backbone grip [22]. Another

model proposes that the binding of the PARP-1 zinc fingers induces a conformational change in the protein which results in PARP-1 activation [23]. Although the PARP-1 zinc fingers are classically known as having a high affinity for nicked DNA, binding with a K_d of 100 nM, they have also been shown to have moderate-to-high affinity for gapped DNA, blunt end DNA, and DNA with 3' extensions [23-25]. A goal of this project is to probe the binding interaction of PARP-like zinc fingers and nicked DNA to provide insight into the structural features of the zinc finger that imparts binding specificity for different forms of damaged DNA.

Due to the ubiquity of zinc finger proteins and their versatility stemming from their ability to recognize specific DNA sequences, zinc finger domains can be important tools for scaffolding in novel protein design. Not only are zinc fingers the most common type of DNA-binding domains found in transcription factors, but they are also widely abundant in the genome, as they are found in approximately 2% of all human genes [26]. One of the main goals in the field of novel zinc finger design is using combinations of zinc fingers to create proteins that can recognize any sequence of DNA [26]. In order to facilitate the rational design of zinc fingers for a particular DNA sequence, Choo has developed a set of recognition rules that summarizes the interactions of specific residues with particular bases in DNA triplets [27]. An objective of rational zinc finger design is to be able to generate zinc fingers that bind to unique sequences in the genome, thus allowing engineered zinc fingers to be used to control gene expression [26]. Such specificity can be achieved by combining up to 6 adjacent zinc fingers in a protein, enabling recognition of an eighteen base pair target [28]. Efforts in gene expression control fuse zinc fingers with activation or repression domains in order to engineer proteins with the ability to regulate endogenous gene expression [29]. In addition, functional domains are being combined with zinc fingers to create novel proteins that bind to specific DNA sequences and have integrase, methylase, or nuclease activities [26]. A long-term application of zinc finger

technology is to ultimately be able to use engineered zinc finger proteins in gene therapy. This approach would be particularly useful for repressing or activating genes that are considered to be not druggable by small molecules and other conventional pharmaceuticals [30]. One potential avenue for such gene therapy is through the use of zinc finger nucleases. Zinc finger nucleases combine sequence-specific zinc fingers coupled to a nuclease domain, allowing the engineered proteins to bind to specific genome targets and cleave specific DNA sites, thus making them extremely useful tools [31]. Various forms of gene editing using zinc finger nucleases are being explored, including gene correction, gene insertion, and gene deletion [30]. The ability of zinc fingers to recognize such a variety of DNA sequences makes them a powerful tool in the future of gene therapy.

One common technique used to design zinc fingers with specific DNA sequence recognition is phage display. Phage display can be described as the artificial evolution of a chemical that binds to a specific target [32] and is a widely used method that allows for the selection of peptides with binding affinity for various targets, including DNA, RNA, or other proteins [33]. By inserting a gene engineered with randomized codons into a phagemid vector and packaging the phage capsid, diverse libraries of randomized proteins can be expressed on the surface of the bacteriophage particles [32]. The most common phage types used are from the M13 and T7 families, both of which involve expressing the randomized protein as a fusion with a coat protein [34]. Phage libraries can accommodate incredible diversity, often consisting of over one billion different peptide sequences. Each phage displays a unique protein from the library, which is physically linked to the DNA that encodes it within the virus capsid [33]. This genotype-phenotype linkage makes the bacteriophage an ideal vessel for propagation of the genetic material encoding for the zinc finger peptide, a crucial feature needed for Darwinian evolution. The phage library is put through rounds of selection, called biopanning, to isolate the

displayed proteins with the highest binding affinity for a target of interest.
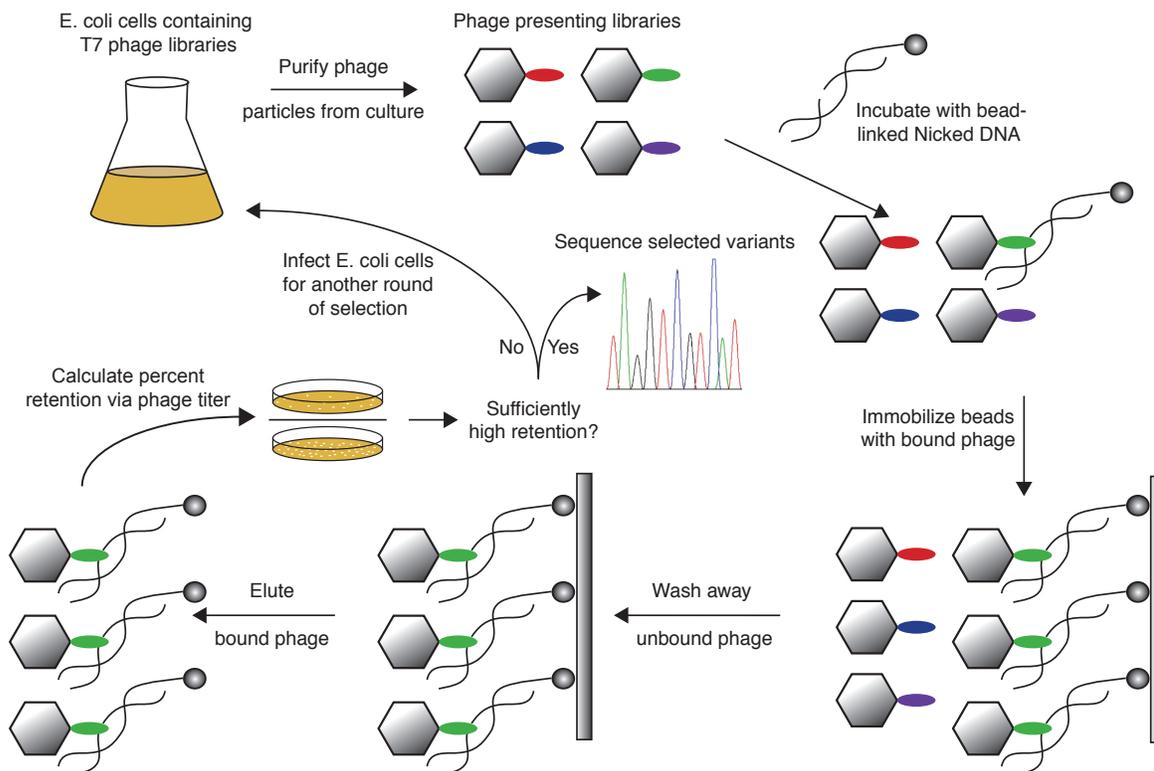
Figure 2. A schematic of the biopanning method that will be used in this project to isolate zinc finger peptides that bind nicked DNA.

The design of the biopanning approach for this project is shown in Figure 2. First, the phage library is incubated with the target, and the phage particles that do not display binding affinity are washed away. Next, the bound phage can be eluted and amplified by infecting a host bacterial strain. Further rounds of selection can take place with higher stringency in order to obtain the peptides with the highest fitness [35]. Factors to alter during biopanning include temperature, salt concentrations, binding and elution time, detergents, and target DNA

concentrations. The genomes of the phage particles that survive selection are sequenced to identify the proteins and additional methods can be carried out to study the selected protein.

Phage display has many useful applications in the biochemical and pharmaceutical realms. In theory, any peptide that possesses a desired property, such as binding affinity to a particular target, can be identified using phage display [33]. Phage display can be applied to clinical diagnostics as well as the creation of novel therapeutics. As an example, phage display has been used to select for peptides that bind to a specific organ or cell type, and these peptides have been used in the creation of targeted drug-delivery systems [36]. Additionally, phage display can be used to create natural protein mimics, called mimitopes, which exert a pharmacological effect on the target; these give insight to the binding properties associated with various hormones, antibodies, enzymes, and growth factors [34]. Mimitopes can be designed as inhibitors of enzymes, thus making them a useful therapeutic approach. A large section of cancer research employs phage display in order to generate proteins that bind to specific structures present on cancer cells but not on regular cells in the hopes of developing targeted therapeutics [33]. The powerful design capabilities of phage display, as well as its wide scope of applications, have made it such a prevalent technology used in current research.

Phage display is also frequently used in the evolution of zinc fingers that bind to novel DNA sequences. Phage display has allowed zinc fingers to be a key tool in conjunction with drug discovery research, as it allows for the selection of zinc fingers that bind to virtually any DNA target sequence [37]. Detailed selection strategies have been outlined that optimize the assembly of up to three zinc fingers in order to recognize unique nine base pair target sites [38]. Phage display has been used to dramatically increase the binding affinity of zinc fingers to a specific target; a 1999 study by Pabo employed phage display technology to increase the binding affinity of the p53 zinc finger to DNA by more than 8-fold [39]. In addition, engineered zinc

10

fingers have been shown to bind to target DNA over 20,000-fold stronger than to nonspecific DNA [38]. A great part of the utility of phage display stems from this ability to confer increase affinity and specificity to displayed peptides.

Although phage display is commonly used to identify zinc finger proteins with binding affinity for novel DNA sequences, the goal of this project is to isolate zinc fingers that bind to structural features of DNA such as DNA nicks or gaps. Selection strategies to identify zinc finger proteins that bind to secondary structures that diverge from canonical B-form DNA have not been yet reported in the literature. We propose a novel strategy to select for PARP-like zinc finger peptides that bind to features of DNA damage. Phage display will be used to force the natural selection of PARP-like zinc finger proteins with strong binding affinity for nicked DNA. Selected peptides will be further studied using various biochemical assays in order to learn more about their structure and binding interaction with nicked DNA. Although this project is currently a proof of principle designed to study the binding interaction of macromolecules, such as proteins and DNA, and knowledge gained from this project could be applied to diagnostics or therapeutics.

Materials and Methods

Materials

The T7 Select® Phage Display System was purchased from Novagen® (Darmstadt, Germany); this included the T7 Select® 10-3b Cloning Kit and the T7 Select® Packaging Kit. These kits contained the T7 Select® phagemid precut with the restriction enzymes EcoRI and HindIII, along with the host E. coli strain BLT5403, T7 phage packaging extracts, and 'UP' and 'DOWN' primers. Custom oligonucleotides used for the phage library construction, biopanning resin construction, and PCR primers were obtained from Operon (Huntsville, AL). Taq DNA Polymerase and all restriction enzymes were purchased from New England Biolabs (Ipswitch, MA). DNA Polymerase I Klenow Fragment (exo–) was obtained from Promega (Madison, WI). The QIAprep Spin Miniprep Kit and QIAquick PCR Purification Kit were purchased from Qiagen (Valencia, CA). The Streptavidin Magnetic Beads for the biopanning resin were obtained from Pierce (Rockford, IL). The BigDye® Terminator v3.1 Cycle Sequencing Kit was purchased from Applied Biosystems (Carlsbad, CA). Centrisep Spin Columns for sequencing purification were obtained from Princeton Separations (Adelphia, NJ), and samples were sequenced using an Applied Biosystems HITACHI 3130 Genetic Analyzer (Grand Island, NY). The pT7-FLAG™-2 Expression Vector, Anti-FLAG® M2 Affinity Gel, FLAG® Peptide, protease inhibitor cocktail, and BL21(DE3)-T1^R Competent Cells were purchased from Sigma® (St. Louis, MO). The Pierce 10% Precise 12-well Tris-Hepes SDS protein gels were purchased from Thermo Fischer Scientific Inc (Rockford, IL).

Construction of the Phage Library DNA Cassette

The phage library was constructed at the DNA level using four custom, single stranded

oligonucleotides. These included eight randomized codons in the DNA cassette for the phage

library, as shown in Table 1.

Table 1. Sequences of oligonucleotides used for phage display library cassette construction. "M" represents adenosine or cytidine, "N" represents any nucleotide, and "K" represents guanosine or thymidine.

1a	5'-GATCCGAATTCAGAATATGCGAAAAGCAACCGCAGCACCTGCAAAGGCTGCATGGAAAAAATTGAAAAAGGCCAGGTGCGCCTGAG-3'
1b	5'-CCGGATGGTACCAGCGATCAATCATGCCMNNMNNCGGMNNMNNCGGATCCACCATTTTTTTGCTCAGGCGCACCTGGC-3'
2a	5'-GATCGCTGGTACCATCCGGGCTGCTTTGTGAAANNKNNKGAAGAACTGNNKNNKCGCCCAGAATATAGCGCGAGCCAGCTGAAAGG-3'
2b	5'-GCCGCAAGCTTCGGCAGCTGTTTTTTCAGCGCTTCTTTATCTTCGGTCGCCAGCAGGCTAAAGCCTTTCAGCTGGCTCGCGC-3'

The oligonucleotides were gel purified using a 16 cm x 18 cm x 1 mm, 8% denaturing

polyacrylamide gel (7 M urea, 1X TBE) run at 450 V for approximately 50 min. The bands were

visualized by UV shadow above a fluorescent TLC plate, and the largest band was excised. The

DNA was extracted from the acrylamide bands overnight in TE buffer (10 mM Tris-HCl, 1 mM

EDTA, pH 8.0) at 4°C while rotating. The supernatant was filtered, concentrated using Amicon

Ultra Centrifugal Filter units (10K MWCO), and ethanol precipitated. The DNA pellets were

resuspended in TE.

Two mutually primed extension reactions were carried out (using oligonucleotides 1a / 1b

and 2a / 2b), in order to create the two halves of the DNA cassette. For each reaction, 0.1 nmol of

each oligonucleotide was slowly annealed in 1x Klenow Buffer in a total volume of 100 μL.

Deoxynucleotide triphosphates (0.0002 M each, final concentration) and Klenow Fragment (9

units/μL final concentration) were added and the reaction was kept at 37°C for 30 min. The

reaction was stopped by isolating the DNA using the Qiagen PCR Purification Kit, then gel

purified as before, but under native conditions.

Up to 25 μg of each duplex was then digested with Kpn1 for one hr at 37°C, and cleaned

up with the Qiagen PCR Purification Kit. In order to minimize spurious homo-duplex ligation, duplex 2 was incubated with calf intestinal phosphatase (CIP) to dephosporylate its 5' ends. The two duplex halves were then ligated using T4 DNA ligase (using a 5-fold molar excess of Duplex 2 relative to Duplex 1 – 61.31 pmol Duplex 1 and 410.7 pmol Duplex 2) at 16°C for 16 hr. The ligated product was gel purified as described for the duplex purification.

To make complementary arms to the precut T7 Select vector, the purified DNA cassette (1.33 µg) was digested separately with EcoRI then HindIII. Each digestion was carried out for 3 hr at 37°C and cleaned up with Qiagen PCR Purification Kit. Ten ng of the digested DNA cassette was ligated into the T7 Select vector arms (using a 3-fold molar excess of insert to vector) with T4 DNA ligase for 16 hr at 16°C. The virus was then packaged *in vitro* by incubating the T7 Select packaging extract along with the vector for 2 hr at 25°C. Ligation efficiency and packaging efficiency were determined using ligation and packaging controls.

Phage Library Amplification and Purification

A plaque assay was carried out to determine the virus titers of the phage containing the desired insert, as well as those of the ligation and packaging controls. Serial dilutions of each sample were prepared, ranging from 10^{-3} to 10^{-7}. For each sample, 10 µL of each phage dilution was combined with 250 µL of log-phase *E. coli* culture and 3 mL of molten top agarose, and plated on LB agar plates containing 50 µg/mL of carbenicillin. The plates were incubated for 4 hr at 37°C. The plaques, which are caused by zones of lysis due to a phage, or plaque-forming unit (pfu), were counted on each plate and multiplied by the dilution value and by 10, to account for the 0.1 ml of dilution plated, thus giving the titer of each phage dilution.

The phage library was amplified using the liquid lysate method. Five hundred mL of LB/carbenicillin media containing 50 µM zinc acetate was inoculated with the *E. coli* strain

BLT5403 and shaken at 37°C until the OD_{600} reached 0.5. At that point, the culture was infected with the phage library with a multiplicity of infection of 0.001, and continued to shake until bacterial lysis was observed. Upon lysis, the culture was centrifuged at 8000 rpm for 10 min to clarify the lysate. The lysate was incubated with 50 μL of DNase I for 30 min to degrade any bacterial DNA. One M NaCl was added to the lysate, incubated on ice for 1 hr, then centrifuged at 8000 rpm for 10 min at 4°C. PEG 8000 was added to the supernatant to a final concentration of 10% w/v. The mixture was centrifuged for 10 min at 8000 rpm and 4°C and the supernatant was discarded. The phage pellet was resuspended in 10% PEG 8000 and centrifuged for 10 min at 7000 rpm and 4°C. The supernatant was removed and the phage pellet was allowed to drain as completely as possible. The phage pellet was resuspended in 10 ml of the zinc-finger buffer, (20 mM Hepes, 50 mM NaCl, 1 mM $MgCl_2$, 50 mM zinc acetate, pH of 7.5) [27]. A plaque assay was carried out to determine the titer of the purified phage stock.

Design and Construction of the Biopanning Resin

The resin for biopanning consists of biotinylated nicked DNA strands bound to streptavidin coated silanized iron oxide magnetic beads. Two different nicked DNA molecules were used for alternating rounds of selection, both 27 base pairs in length, along with their corresponding non-nicked, non-biotinylated counterparts to be used in the mobile phase.

Table 2. Sequences of the biotinylated strand of the nick containing duplexes used for biopanning. The apostrophe shows the location of the nick on the complementary strands.

Nick 1	5'-biotin-GGATCGGTGGATGCAAG'CTATGCTACG-3'
Nick 2	5'-biotin-GCTGAGTGGGTAACGCG'ATCTATGACG-3'

The nicked construct was composed of a biotinylated 27 bp oligonucleotide annealed to two complimentary 17 and 10 bp oligos, with the nick between the 17th and 18th bases. Before annealing, the 5' hydroxyl-terminus of each 17-mer was phosphorylated using T4 polynucleotide

kinase (300 pmol of oligonucleotides phosphorylated using 10 units of enzyme at 37°C for 30 min). The 17-mer and 9-mer were slowly annealed in 1.5 times molar excess to the biotinylated complimentary 27-mer (148 ng 27-mer, 527 ng 17-mer, 874 ng 9-mer) and then added to streptavidin-coated beads in 100-fold excess molar of biotin binding capacity of the beads (8 nmol biotin/mL bead). Free competitive duplex DNA strands, with the same sequence as two nicked strands, were created by annealing the two complementary 27 bp oligonucleotides in equimolar amounts. To test the efficiency of the annealing process, as well as the binding of the biotinylated nicked DNA strands to the streptavidin beads, samples from each step were separated on a 20% native polyacrylamide gel run at 200 V for 1 hr.

PCR Confirmation of DNA Insert

The presence of the DNA insert in the phage vector was confirmed via PCR amplification. Of the original phage stock, approximately 6×10^{10} phage particles were digested with 100 μg/mL proteinase K and 0.5% SDS in a total volume of 100 μL for 2 hr at 56°C. The viral DNA was extracted from the reaction mixture with an equal volume of phenol, pH 7.9. Extraction was repeated twice on the aqueous phase using a 1:1 mixture of phenol:chloroform, then phage DNA purified using ethanol precipitation, as described previously. PCR was performed using ThermoPol Taq DNA polymerase and the primers provided in the T7 Select kit on 27.6 ng of the extracted phage DNA according to vendor protocol. The PCR reaction was cleaned up using the Qiagen PCR Purification Kit and the products were run on a 1 mm, 8% native polyacrylamide gel for 40 minutes at 150 V.

Biopanning for Peptides that Bind Nicked DNA

Four rounds of biopanning were carried out using two alternating biotinylated nicked DNA constructs as previous described, denoted N1 and N2, as well as their corresponding free non-nicked counterparts, D1 and D2. Prior to each round of selection, 50 μL of magnetic beads were prepared by washing once with the zinc-finger buffer, using a magnet to aggregate the beads. The beads were then incubated with the desired nicked molecule for 10 min at 25°C, with 100-fold molar excess of biotinylated DNA to streptavidin binding sites. Each set of magnetic beads was then blocked with 100 μL of blocking buffer (20 mM Hepes, 50 mM NaCl, 1 mM MgCl$_2$, 50 mM zinc acetate, 6% fat-free dry milk, pH 7.5) for 1 hr at 25°C. For each round of selection, 10^{11} phage particles were applied to the prepared magnetic beads, along with the corresponding competitor DNA duplex (free duplex DNA at one tenth the molar concentration of biotinylated nicked DNA). The volume was brought to 1.5 mL in binding buffer (20 mM Hepes, 50 mM NaCl, 1 mM MgCl$_2$, 50 mM zinc acetate, 2% fat-free dry milk, 1% Tween 20, pH 7.5) and rotated during the binding reaction. The four rounds of selection were as follows: Round 1 – N1 for 1 h at 4°C, Round 2 – N2 for 1 h at 4°C, Round 3 – N1 for 1 h at 25°C, and Round 4 – N2 for 1 h at 25°C. Following each round, the beads were washed five times with the zinc-finger buffer to remove all unbound phage. The supernatant from each wash was combined and a plaque assay was performed on the supernatant to calculate the percent retention of that round of biopanning. The beads were shaken in 100 μL of 0.1 M triethylamine for five min to elute the bound phage. The beads were then neutralized with an equal volume of Tris-HCl, pH 7.5, and added to 50 mL of log-phase *E. coli* culture to amplify the elute phage. The amplified phage were purified by PEG precipitation as previously described. The resulting phage stock was quantified via a plaque assay before being applied to the next round.

Sequencing of Phage Surviving Biopanning

Following the four rounds of biopanning, 20 phage plaques were selected for sequencing. Plaques were picked and added to 1 mL of log-phase *E. coli* culture. Following shaking and lysis, the phagemid DNA was isolated using the Qiagen Miniprep kit. The presence of the insert was confirmed via PCR, as described previously, and the products were visualized on a 1% agarose gel, run at 140 V for 1 hr. The phagemid DNA was sequenced using Applied Biosystem's BigDye Terminator v3.1 Cycle Sequencing Kit. The PCR Sequencing reaction using phagemid DNA contained the following: 3 μL 5X Sequencing Buffer, 0.5 μL T7 Up Primer, 4 μL phage DNA, 1.5 μL dH20, and 1 μL BigDye 3.1. The thermocycler conditions were: 95°C for 2 min, 40 cycles of 95°C for 10 sec, 50°C for 5 sec, and 60°C for 4 min. The sequencing reactions were purified with Centri-Sep columns and sequenced using the ABI Prism 3130 Genetic Analyzer.

Cloning of Selected Gene into E. Coli for Expression and Purification

In those phage which had the desired peptide-insert sequence, PCR was performed to amplify the insert and add additional features to the gene for cloning, using the following primers.

Table 3. Sequences of the primers used to amplify the desired sequence for cloning. The HindIII restriction site is shown in blue, FLAG tag in orange, N-terminal cysteine in green, and EcoRI restriction site in red.

Forward Primer	5'-CACGTCAAGCTTGATTATAAAGATGATGATGATAAATGCGGCG GCGAATATGCG AAAAGCAACCGC-3'
Reverse Primer	5'-GTCAGGGAATTCTCAATCATGCCATCCAGCGGG-3'

Standard PCR was carried out using 0.5 μL phage DNA. The thermocycler conditions were: 95°C for 5 min, 35 cycles of 95°C for 30 sec, 62°C for 45 sec, and 68°C for 30 sec, then 68°C for 5 min. The PCR reaction was cleaned up using the Qiagen PCR Purification Kit and the

amplicon size was verified on a 1% agarose gel.

To create complementary ends for cloning, both the PCR product insert (0.5 µg) and the pT7-FLAG-2 Expression Vector (0.2 µg) were restriction digested separately with EcoRI then HindIII. Each digestion was carried out for 1 hr at 37°C, and each were cleaned up with Qiagen PCR Purification Kit. The ligation reaction was carried out with a vector:insert ratio of 1:3 (50 ng vector, 150 ng insert) with T4 DNA Ligase at 25°C for 10 minutes, then added directly to BL21(DE3)-T1^R competent cells for heat shock transformation. The competent cells were first incubated on ice for 30 min, then moved to at 37°C water bath for 45 sec, then incubated again on ice for 2 min. 450 µL of SOC media was added to the cells and they were shaken at 225 rpm at 37°C for 1 hr. 100 µL of the bacterial culture was plated onto pre-warmed LB-agar plates containing 100 µg/mL of ampicillin, and incubated overnight at 37°C. Colonies were picked to inoculate 3 mL LB cultures, which were grown until log-phase, then the plasmid was isolated using the Qiagen Miniprep Kit. To verify the presence of the insert in the vector, plasmid DNA was digested with BamHI, a unique restriction site in the insert, for 1 hr at 37°C. The cut and uncut plasmids were run on a 1% agarose gel for 2 h at 145V. Selected colonies were also sequenced as previously described in order to confirm the presence of the gene insert.

Peptide Expression and Purification

A 50 mL log-phase *E. coli* culture containing the pT7-FLAG-2-insert plasmid was induced to express the selected peptide, pGuerra-1, with 1.0 mM IPTG while shaking for 3 hr at 37°C. The cells were harvested, frozen overnight at -20°C, and resuspended in 10 mL of lysis buffer (50 mM Tris-HCl, pH 7.4, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100, protease inhibitor cocktail) per gram of wet cells. The cells were lysed mechanically, via vigorous pipetting and stirring, and the lysate was clarified by centrifugation at 21,000 *x g* for 15 min at

19

4°C. The crude cell lysate was added to anti-FLAG M2 affinity agarose beads, pre-equilibrated

in TBS (50 mM Tris-HCl, pH 7.4, 150 mM NaCl), and incubated for 4 hours at 4°C while

rotating. The resin was the spun down at 8200 $x\ g$ for 1 min and washed 3 times with cold TBS.

The peptide was eluted with 100 μL of FLAG elution solution (150 ng/μl FLAG peptide, 0.5 M

Tris-HCl, pH 7.5, 1 M NaCl) for 30 min while rotating. Protein aliquots from the crude lysate,

each wash, and eluted peptide were run on a denaturing SDS-PAGE to assess the purification.

Results

The randomized zinc finger DNA cassette was constructed with a modest yield

A phage library of randomized zinc finger peptides was constructed at the DNA level using synthetic oligonucleotides with built in randomized codons using the NNK scheme. The PAGE-purified pairs of oligonucleotides were annealed and extended using Klenow fragment to yield the two halves of the DNA cassette (Figure 3).
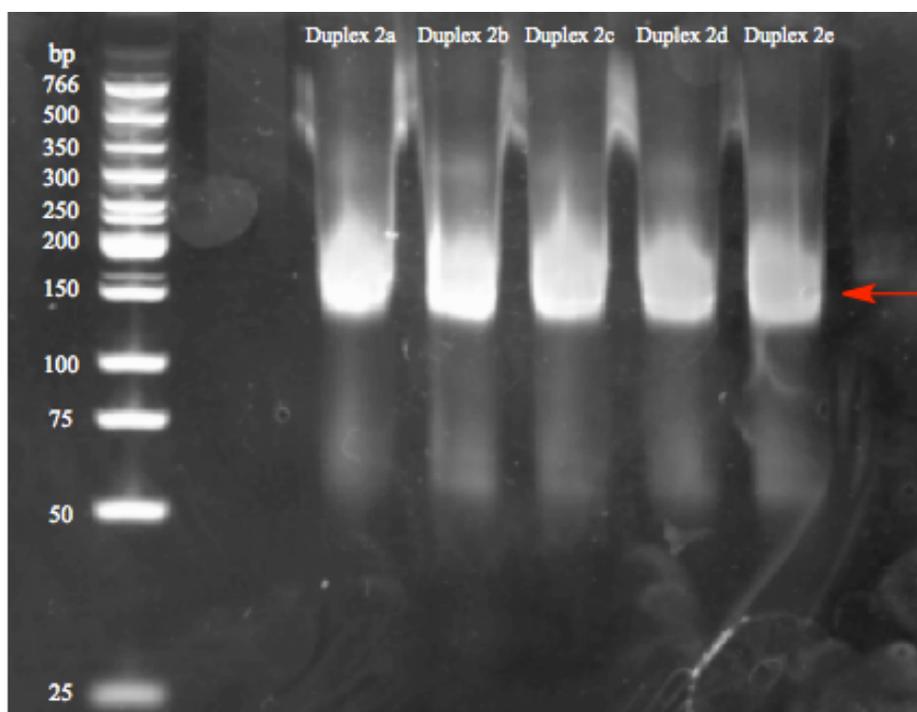
Figure 3. An ethidium-bromide stained polyacrylamide gel of the mutually primed extension reaction producing Duplex 2. The large bands at 150 bp correspond to the desired products.

One of the likely challenges with the ligation of the two duplex halves is the spurious ligation of like halves via the KpnI-cut ends. In order to increase the yield of the desired product, the halves were ligated in a ratio of 1:5 of Duplex 1 to CIP-treated Duplex 2. The CIP treated duplex will not be able to self-ligate due to its dephosphorylated end, so the goal was that the

excess of the CIP treated Duplex 2 would outcompete the self-ligation of Duplex 1. The purification gel of this ligation product is shown in Figure 4.
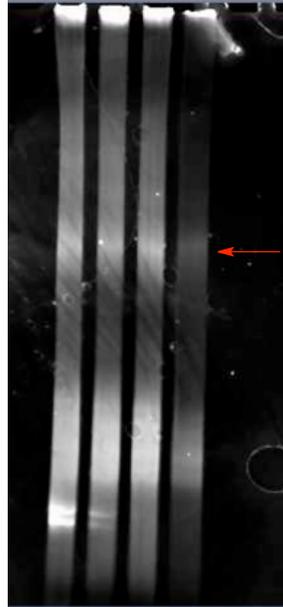
Figure 4. An ethidium-bromide stained native polyacrylamide gel of the ligation reaction. The presumed ligation product corresponds to the predicted 300 bp arrow. There is also a large amount of un-ligated duplex pieces further down on the gel, resulting primarily from the excess of the CIP-treated Duplex 2.

The ligation reaction and purification resulted in a 23.6% yield as measured via the OD_{600} with the Nanodrop. Following the two enzyme digestions with EcoRI and HindIII, the final amount of the DNA cassette was 621 ng, an overall yield of 11.0%.

Final phage library contained sufficient diversity for biopanning and presence of DNA cassette was confirmed in phagemid

After the ligation of the DNA cassette into the phage vector and the packaging of the vector, the resulting titer of the phage library was 1.4×10^{10} pfu/ml, which indicated that the library contained sufficient diversity for biopanning. The phage library was designed with eight

randomized codons, thus the total possible unique variants would be 2.56 x 10^{10}. Although the phage titer indicated the library did not contain the total possible number of variants, it still encompassed enough variation for selection. The liquid lysate amplification method was the preferred method for a library of over 10^6 recombinants, so the library was amplified using *E. coli* cultured in LB broth containing zinc acetate to allow the zinc finger proteins on the surface of the phage to coordinate with the zinc ions. After purifying the amplified library, the final titer was 6.7 x 10^{11} pfu/ml, thus the 10 ml stock contained enough phage particles to perform approximately 66 rounds of biopanning, using 10^{11} phage particles for each round. The presence on the DNA cassette in the phage was confirmed using PCR. Proteinase K digestion of phage particles was necessary to yield enough phage DNA for PCR. The PCR products were separated on a polyacrylamide gel, shown in Figure 5.
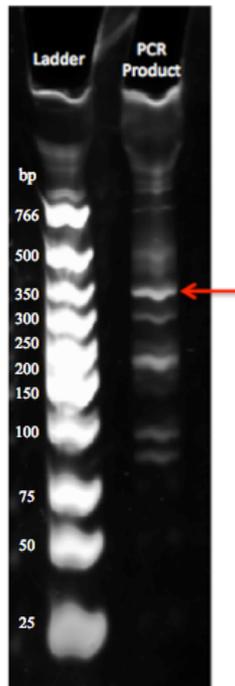
Figure 5. A GelStar stained native polyacrylamide gel used to confirm the presence of the DNA cassette insert in the PCR amplified phage DNA. The distinct band just above the 350 bp marker corresponds to the insert in the phage vector (actual size of 368 bp).

Nicked duplex for biopanning resin annealed and bound to magnetic beads

The resin for biopanning was constructed using custom oligonucleotides to form a biotinylated nicked DNA duplex of 27 base pairs, which bound to magnetic beads coated with streptavidin. A 27-base biotinylated oligonucleotide was annealed with two complementary oligonucleotides to create the nicked duplex. The test gel in Figure 6 confirms that the duplex annealed and was bound to the magnetic beads.
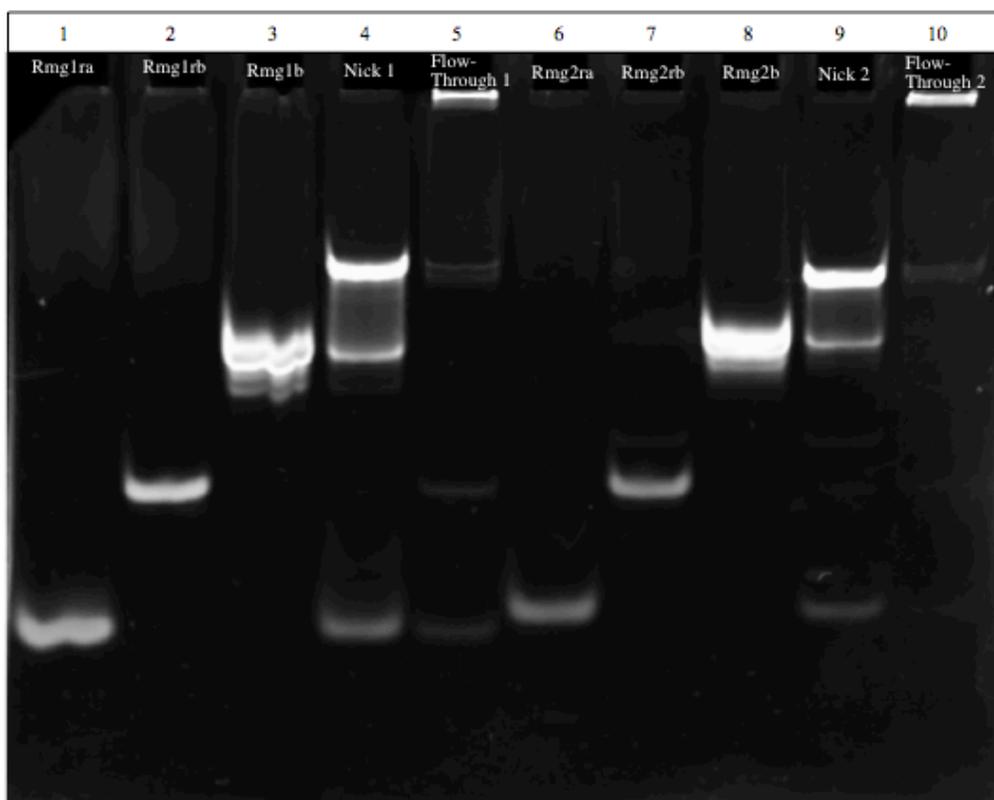
Figure 6. A GelStar stained native polyacrylamide gel used to determine if the nicked strands had annealed correctly and also bound to the magnetic beads. Lanes 1 and 6 contain the 10-mers, Rmg1ra and Rmg2ra. Lanes 2 and 7 contain the 17-mers, Rmg1rb and Rmg2rb. Lanes 3 and 8 contain the biotinylated 27-mers, Rmg1b and Rmg2b. Lanes 4 and 9 contain the annealed 7-, 10, and 27-mers to create the nicked strands, Nick 1 and Nick 2. Lanes 5 and 10 contain the supernatant of the respective nicks incubated with the streptavidin-coated magnetic beads.

Four rounds of biopanning selected for a single variant from the phage library

Four rounds of selection were performed to enrich the phage library for phage-displayed peptides that bind with high affinity to nicked DNA. The two nicked sequences, N1 and N2, were used in alternating rounds, along with their corresponding soluble non-nicked sequence. Rounds 1 and 2 were carried out at 4°C for 1 hr, while rounds 3 and 4 were at 25°C for 1 hr. Phage retentions from rounds 1 and 2 were unquantifiable by means of plaque assay (Figure 7). However, there was sufficient eluted phage from both rounds to amplify. In rounds 3 and 4, retention was measured at 43.2% and 62%, respectively.
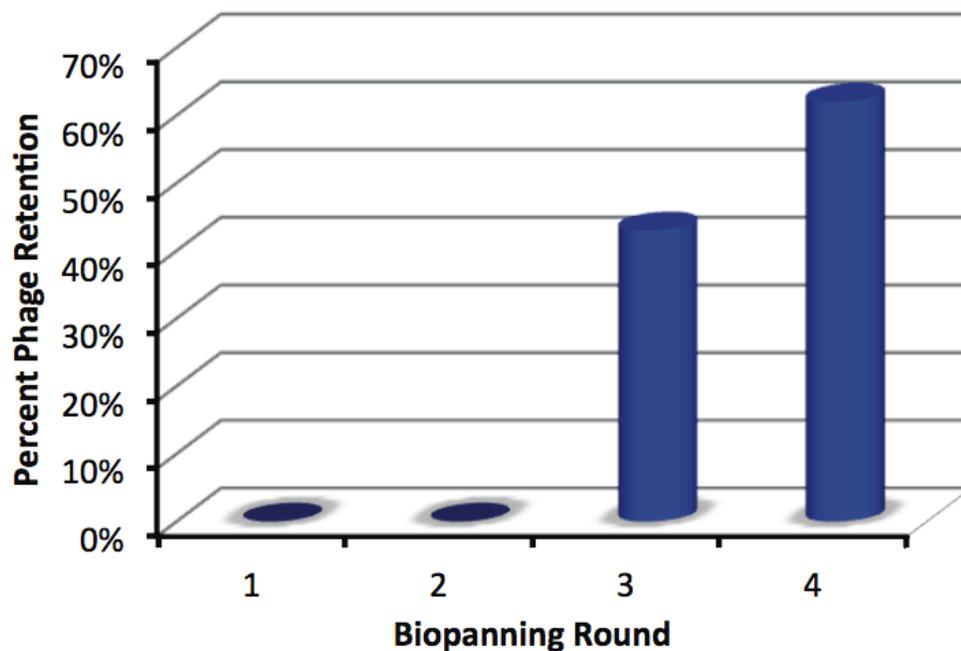
Figure 7. Percent of phage particles retained on beads following each round of selection.

Thus in the final two rounds, significant enrichment of the phage library was observed, and 20 phage isolates from the final round were sequenced. Of these, 17 were successfully sequenced. The sequences of the all 17 phage were identical. A representative sequence is shown in Figure 8.

```
              10        20        30        40        50        60        70        80
QUERY   GATCCGAATTCAGAATATGCGAAAAGCAACCGCAGCACCTGCAAAGGCTGCATGGAAAAAATTGAAAAAGGCCAGGTGCG
        :::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
QUERY   GATCCGAATTCAGAATATGCGAAAAGCAACCGCAGCACCTGCAAAGGCTGCATGGAAAAAATTGAAAAAGGCCAGGTGCG
              10        20        30        40        50        60        70        80


              90       100       110       120       130       140       150
QUERY   CCTGAGCAAAAAAATGGTGGATCCGNNKNNK-CCGNNKNNKGGCATGATTGATCGCTGGTACCATCCGGGCTGCTTTGTG
        :::::::::::::::::::::::::::   .   . :::   .  .::::::::::::::::::::::::::::::       ::
QUERY   CCTGAGCAAAAAAATGGTGGATCCGTATACGCCCGCTGGATGGCATGATTGATCGCTGGTACCATCCGGGC-------TG
              90       100       110       120       130       140       150


           160       170       180       190       200       210       220       230
QUERY   AAAANNKNNKGAAGAACTGNNKNNKCGCCCAGAATATAGCGCGAGCCAGCTGAAAGGCTTTAGCCTGCTGGCGACCGAAGA
        :::   .   .::::::::::   .   .:::::::::::::::::::::::::::::::::::::::::::::::::::::::::
QUERY   AAAGGGGGGGGAAGAACTGGTGTGTCGCCCAGAATATAGCGCGAGCCAGCTGAAAGGCTTTAGCCTGCTGGCGACCGAAGA
           160       170       180       190       200       210       220       230


           240       250       260       270
QUERY   TAAAGAAGCGCTGAAAAAACAGCTGCCGAAGCTTGCGGC
        :::::::::::::::::::::::::::::: :::::::
QUERY   TAAAGAAGCGCTGAAAAAACAGCTGCCGAAGC-TGCGGC
           240       250       260       270
```

Figure 8. The alignment of a phage insert sequence (bottom line) with the sequence of the original DNA construct (top line). ':' denotes identical matches and '.' denotes conservative replacements.

Although the sequenced phage insert contained 88.2% identity to the original construct, some major differences were observed. A single nucleotide insertion of a cytosine after bp 112 in the sequenced DNA shifts the reading frame for the remainder of the insert. In addition, this shift results in a stop codon at bp 130, resulting in the truncation of the peptide. It is also important to note that the sequenced DNA displays a 7 base pair deletion following bp 153. The sequences of all 17 phage are shown in Appendix A.

The selected DNA sequence was cloned into E. coli for expression and purification

In order to express the selected variant in *E. coli*, the corresponding DNA was amplified from selected plaques by PCR and confirmed by gel electrophoresis (Figure 9).

26

Figure 9. GelStar stained agarose gel to confirm size of PCR amplicon prior to cloning. Each lane, representing the amplicon of a picked phage plaque, has one distinct band slightly below the 200 bp marker (actual size 172 bp).

Following cloning of the insert into the expression vector and heat shock transformation, colonies were selected and the plasmids were isolated. To confirm that these colonies had plasmids containing the desired insert, the plasmids were digested with BamHI, for which there is only one site in the middle of the insert, and verified on an agarose gel (Figure 10).
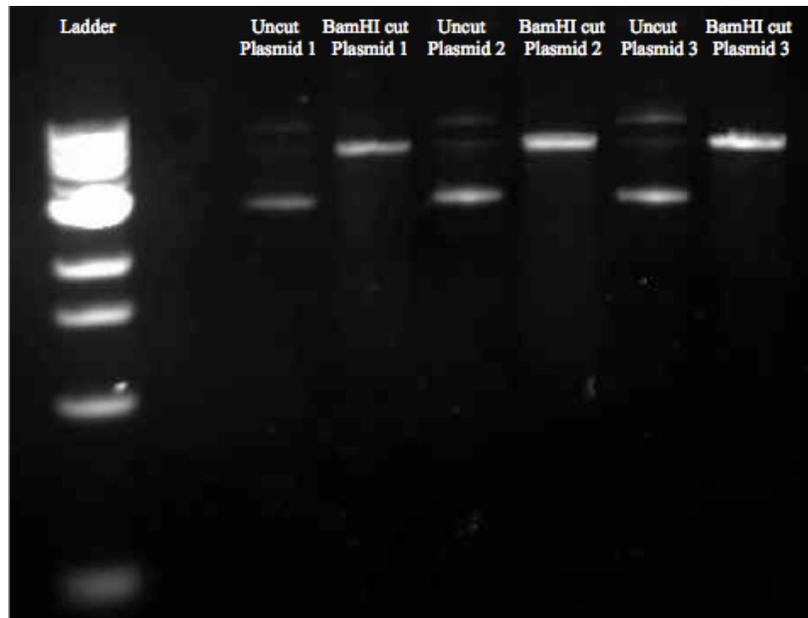
Figure 10. GelStar stained agarose gel of three sample plasmids uncut and cut with BamHI. Lanes 1, 3, and 5 have the characteristic banding pattern of an uncut plasmid, where as Lanes 2, 4, and 6 display one band corresponding to the linearize plasmid (4.8 kb).

Following cloning of the expression vector into appropriate competent cells, protein expression was induced with IPTG. Samples were taken every 30 min following induction and assessed for protein content on a 4-20% gradient SDS-PAGE gel (Figure 11). Due to the small size of the protein (5.6 kDa), it was difficult to determine if the protein was in fact being expressed. As an alternative to electrophoresis, mass spectrometry will be used following affinity purification to determine if the protein of the desired size was purified.
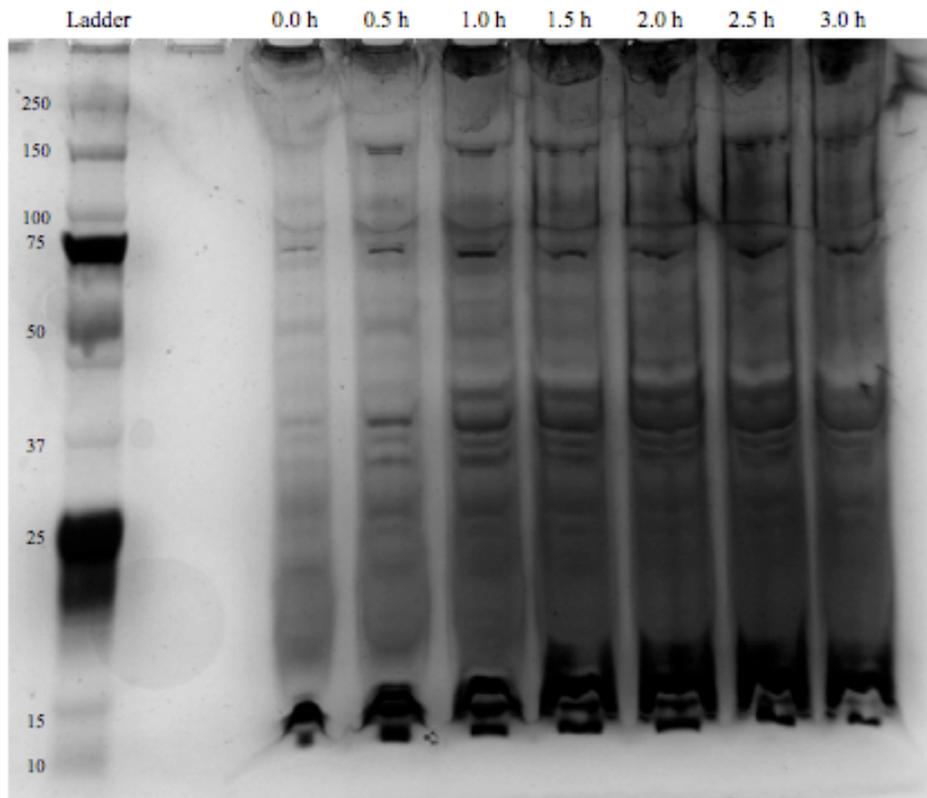
Figure 11. Silver stained 4-20% SDS-PAGE of crude protein samples at 30 min time points following induction of expression with IPTG.

Discussion

The goal of this project was to design a novel strategy for the selection of proteins that bind to structural features of DNA, a technique that has never been reported, while simultaneously studying the binding interaction of zinc finger proteins and nicked DNA. The second zinc finger of DNA repair enzyme poly(ADP-ribose) polymerase-1, or PARP-1, was selected as the template for the phage library as it is known for binding to DNA nicks. The DNA and amino acid sequence of the DNA cassette is shown in Figure 12. To introduce sufficient diversity in the phage library, eight codons were randomized in the zinc finger peptide (shown in red), where "n" is any nucleotide and "k" is either thymidine or guanidine. This allows for all 20 amino acids to be encoded, while at the same time eliminating two stop codons.

```
              E   Y   A   K   S   N   R   S   T   C   K   G   C   M
5'-gat ccg aat tca gaa tat gcg aaa agc aac cgc agc acc tgc aaa ggc tgc atg
        EcoRI

     E   K   I   E   K   G   Q   V   R   L   S   K   K   M   V   D   P   X
   gaa aaa att gaa aaa ggc cag gtg cgc ctg agc aaa aaa atg gtg gat ccg nnk

    X   P   X   X   G   M   I   D   R   W   Y   H   P   G   C   F   V   K   X
  nnk ccg nnk nnk ggc atg att gat cgc tgg tac cat ccg ggc tgc ttt gtg aaa xxx
                                       KpnI

   X   E   E   L   X   X   R   P   E   Y   S   A   S   Q   L   K   G   F   S
 nnk gaa gaa ctg nnk nnk cgc cca gaa tat agc gcg agc cag ctg aaa ggc ttt agc

   L   L   A   T   E   D   K   E   A   L   K   K   Q   L   P
 ctg ctg gcg acc gaa gat aaa gaa gcg ctg aaa aaa cag ctg ccg aag ctt gcg gc-3'
                                                    HindIII
```

Figure 12. Sequence of the phage library; the DNA code is shown in lower case and the amino acid code shown in upper case.

The codons that are randomized comprise part of the loop regions of the zinc finger protein, which are the regions that are responsible for binding to DNA [24]. The goal of the phage display process is to identify a unique PARP-like zinc finger protein with increased binding affinity to nicked DNA.

The DNA cassette contains three specific restriction sites, which are underlined in Figure 12. The Kpn1 site allows for the two duplex halves to be ligated together, creating the full-length cassette. It was necessary for the cassette to be made in two halves, as custom-designed synthetic oligonucleotides are only commercially available up to 100 bases in length. The four initial single-stranded oligonucleotides form two pairs, which contain an overlapping region of 20 base pairs. The single-stranded oligonucleotide pairs are annealed and extended to form two duplex halves using Klenow Fragment in a mutually primed extension reaction; the schematic of this process is depicted in Figure 13. Additionally, the cassette contains an EcoRI site on one end and a HindIII site on the other end in order to insert the cassette into the phage vector (Novagen T7 Select System), which contains arms precut with EcoRI and HindIII. After ligating the DNA cassette into the phagemid vector, it can be packaged spontaneously using T7 capsid protein extracts, thus creating the phage library.
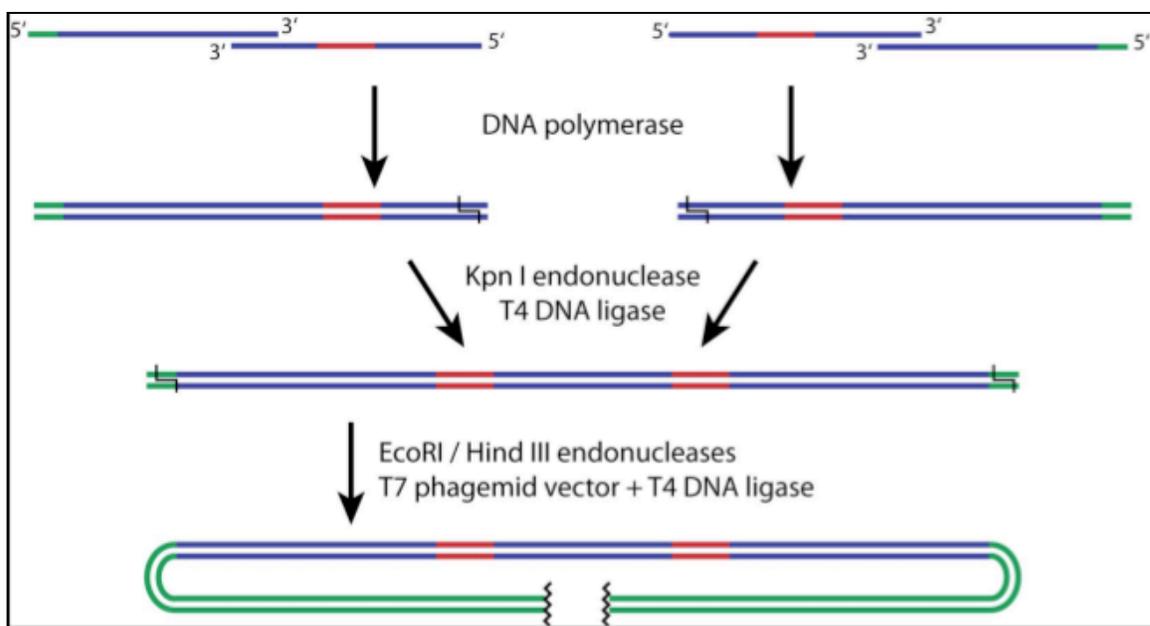
Figure 13. The scheme of the DNA cassette construction.

The molecular biology of the construction of the DNA cassette presented many challenges that were eventually overcome in the creation of the phage library. One consistent problem was a loss in yield during each round of DNA purification with polyacrylamide gels. Although the crush and soak method had increased the yield from previous attempts at the library construction, product was still lost each time during this step. Additionally, another large issue was the palindromic restriction site of Kpn1 and the tendency of the duplex halves to self-ligate. In order to prevent this, the duplex halves were ligated together with a great excess of the CIP-treated Duplex 2. Though the yield from this ligation reaction was not much greater that past yields, it most likely contained greater amounts of the full length DNA cassette and less self-ligated strands. The ligation of the DNA cassette into the phage vector, packaging of the phage particles, and plaque assays confirmed that there was a sufficient amount of the DNA cassette, and thus it resulted in a phage library of the desired size for biopanning. The amplification and

purification of the phage library has allowed for the creation of ample stocks for biopanning in the future.

Short strands of nicked DNA were chosen as the target for biopanning in order to first select for nick-binding zinc fingers with the highest affinity. PARP-1 binds only to DNA structures such as nicks and not specific DNA sequences, thus two different random nicked sequences were chosen as targets for selection. Each strand is 27 base pairs in length and contains a nick between the 10th and 11th base pairs. Silanized iron magnetic beads coated with streptavidin were chosen as the solid resin for the biotinylated nicked DNA to provide a method for easy separation of the nicked DNA strands with bound phage. A magnet can then be used to separate the beads and bound phage particles from the soluble unbound phage particles during rounds of selection, and the bound phage can be eluted from the DNA using a strong base, triethylamine. Multiple rounds of biopanning were carried out to isolate phage particles with the highest binding affinity, and various levels of control were employed to ensure that only phage that bind nicked DNA were selected. Separate rounds were conducted alternating each nicked strand bound to the beads; the two different nicked sequences (N1 and N2) were used in order to prevent the binding of any sequence-specific zinc fingers. In addition, soluble non-nicked DNA strands with the same sequence as the nicked strands were used as competitor DNA during rounds of selection to eliminate any zinc fingers that bind nonspecifically to DNA. In the four rounds of biopanning, temperature was the main factor altered to increase stringency; the first two rounds were performed at 4°C and the second two at 25°C. Increasing the binding temperature results in a shift of the K_d, thus favoring peptides that bind tighter and increasing the stringency of selection. Increasing the temperature also allows for the selection structured peptides that bind to nicked DNA in more physiologically relevant conditions. Other factors that were considered to increase the stringency of the selection included decreasing the binding time,

decreasing the eluting time, and increasing the salt concentration. However, the four rounds of selection led to sufficient enrichment of the phage library, so sequencing was then pursued.

Out of the enriched phage pool at the end of biopanning round 4, the 17 selected sequences were identical, which would suggest that the biopanning was successful in directing the *in vitro* evolution of the phage to a selective pool of phage with increased binding affinity for nicked DNA. The selected sequence, however, contained one inserted base pair in each phage sequence which shifts the reading frame of the remaining peptide and introduces a stop codon. This stop codon results in the premature termination of the peptide, and thus the expressed peptide was only 39 residues. These N-terminal residues comprise the β-sheet regions of the PARP zinc finger, which is the region responsible for providing the binding interface with DNA. Due to the frameshift and premature stop codon, only two of the selected randomized codons were incorporated into the resulting peptide. The first randomized codon, which was glutamate in the original zinc finger, was found to be tyrosine in the sequenced phage. The second randomized codon, originally lysine, was threonine in the sequenced phage. Both residues contain hydroxyl groups, which could stabilize the peptide's interaction with DNA through hydrogen bonding with the polar backbone of DNA. Although it is very plausible that the selected peptide is merely an artifact of selection and does retain any ability to bind DNA, it is possible that is does have some affinity for nicked DNA. The selected peptide was therefore cloned into *E. coli* for expression and purification in order to carry out binding studies to determine whether it does in fact have affinity for nicked DNA.

The gene for the selected peptide was cloned into the pT7-FLAG-2 expression vector with an N-terminal FLAG-tag to allow the peptide to be purified with an anti-FLAG affinity resin. In addition, an N-terminal cysteine residue was added to the peptide to allow for chemoselective labeling with a fluorophore for fluorescence polarization binding assays with

DNA. In a scheme developed by Schuler *et al.*, the N-terminal cysteine of a peptide is selectively

reacted with a thioester fluorophore via a native chemical ligation, as seen in Figure 14 [40].
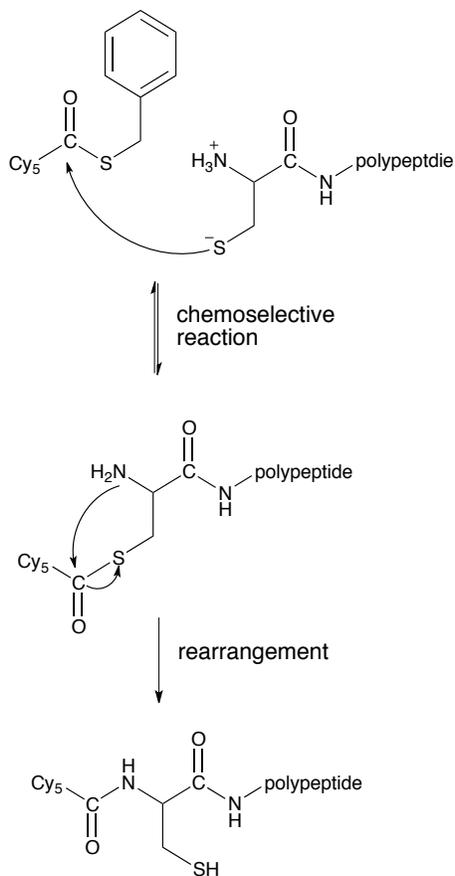
Figure 14. Chemoselective labeling of N-terminal cysteine with thioester fluorophore [40]

As fluorophores containing thioesters are not commercially available, a Cy5-thioester dye will be

synthesized in a two-step reaction developed by Schuler *et al*. Benzyl mercaptan will first be

activated with trimethylaluminum, then reacted with Cy5-succinimidyl ester in order to form the

desired Cy5-thioester [40]. Following fluorophore labeling of the peptide, fluorescence

polarization assays will be performed to measure its binding to nicked DNA, as well as to intact

DNA and gapped DNA. These binding assays will provide insight as to whether the peptide does

in fact retain any binding affinity or specificity for nicked DNA.

In addition to studying the binding interaction of zinc fingers and nicked DNA, another goal of this project is to generate a molecule with potentially useful diagnostic or therapeutic applications. A small fluorophore-conjugated peptide with the ability to bind with high affinity to nicked DNA could serve as an indicator of nicked DNA *in vivo*. This could provide diagnostic information about cells' genomic instability, which is an important hallmark of cancer, as well as other diseases. A small peptide generated in this study could also be applied to cancer therapeutics, as many cancer therapies target inhibiting DNA repair. Potentially, this peptide could bind selectively and tightly to nicked DNA, thus preventing PARP and other DNA repair enzymes from recognizing and repairing the lesion. The inhibition of DNA repair results in cell cycle arrest and eventually cell death via apoptosis. A peptide identified from this study could act analogously to PARP inhibitors, which have emerged as promising avenues of therapeutics in cancer cells deficient in BRCA1 and BRCA2 [41]. By binding to nicked DNA and obstructing DNA repair, this peptide could potentially sensitize cancer cells to other chemotherapeutic agents.

This study reports a novel selection method of isolating peptides that bind to structural features rather than specific sequences of DNA. Four round of biopanning of PARP-like zinc finger peptides that bind to nicked DNA resulted in the convergence of a diverse phage library to a single unique sequence. Interestingly, the selected sequence contained a one base pair insertion approximately half way through the sequence, resulting in a frame shift and the introduction of a stop codon. This truncated peptide comprises the loop and β-sheet regions of the original zinc finger, which were surmised to be regions critical in DNA binding, thus it is possible that the selected peptide does have affinity for nicked DNA. Further work needs to be carried out to ascertain if this selected peptide actually retains binding affinity and specificity for nicked DNA, or whether it is just an artifact of selection.

In the future, the phage library will be used to select for zinc fingers that bind to other structural features of DNA, such as gapped DNA and abasic DNA. The comparison of the resulting peptides could provide valuable insight to the structural and sequence features of zinc fingers that allow for macromolecular recognition and differentiation of various structural features of DNA. Aside from demonstrating a proof of principle, it is possible that this novel selection strategy could give rise to peptides with useful diagnostic and therapeutic applications.
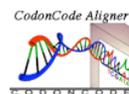
```
<< RMG004D_...  GATCC GAATT CAGAA TATGC GAAAA GCAAC CGCAG CACCT GCAAA GGCTG CATGG AAAAA ATTGA AAAAG
<< RMG013D_...  GATCC GAATT CAGAA TATGC GAAAA GCAAC CGCAG CACCT GCAAA GGCTG CATGG AAAAA ATTGA AAAAG
<< RMG008D_...  GATCC GAATT CAGAA TATGC GAAAA GCAAC CGCAG CACCT GCAAA GGCTG CATGG AAAAA ATTGA AAAAG
<< RMG015D_...  GATCC GAATT CAGAA TATGC GAAAA GCAAC CGCAG CACCT GCAAA GGCTG CATGG AAAAA ATTGA AAAAG
<< RMG012D_...  GATCC GAATT CAGAA TATGC GAAAA GCAAC CGCAG CACCT GCAAA GGCTG CATGG AAAAA ATTGA AAAAG
<< RMG018D_...  GATCC GAATT CAGAA TATGC GAAAA GCAAC CGCAG CACCT GCAAA GGCTG CATGG AAAAA ATTGA AAAAG
<< RMG003D_...  GATCC GAATT CAGAA TATGC GAAAA GCAAC CGCAG CACCT GCAAA GGCTG CATGG AAAAA ATTGA AAAAG
<< RMG016D_...  GATCC GAATT CAGAA TATGC GAAAA GCAAC CGCAG CACCT GCAAA GGCTG CATGG AAAAA ATTGA AAAAG
<< RMG019D_...  GATCC GAATT CAGAA TATGC GAAAA GCAAC CGCAG CACCT GCAAA GGCTG CATGG AAAAA ATTGA AAAAG
<< RMG017D_...  GATCC GAATT CAGAA TATGC GAAAA GCAAC CGCAG CACCT GCAAA GGCTG CATGG AAAAA ATTGA AAAAG
<< RMG009D_...  GATCC GAATT CAGAA TATGC GAAAA GCAAC CGCAG CACCT GCAAA GGCTG CATGG AAAAA ATTGA AAAAG
<< RMG007D_...  GATCC GAATT CAGAA TATGC GAAAA GCAAC CGCAG CACCT GCAAA GGCTG CATGG AAAAA ATTGA AAAAG
<< RMG002D_...  GATCC GAATT CAGAA TATGC GAAAA GCAAC CGCAG CACCT GCAAA GGCTG CATGG AAAAA ATTGA AAAAG
<< RMG005D_...  GATCC GAATT CAGAA TATGC GAAAA GCAAC CGCAG CACCT GCAAA GGCTG CATGG AAAAA ATTGA AAAAG
<< RMG010D_...  GATCC GAATT CAGAA TATGC GAAAA GCAAC CGCAG CACCT GCAAA GGCTG CATGG AAAAA ATTGA AAAAG
<< RMG022D_...  GATCC GAATT CAGAA TATGC GAAAA GCAAC CGCAG CACCT GCAAA GGCTG CATGG AAAAA ATTGA AAAAG
<< RMG011D_...  GATCC GAATT CAGAA TATGC GAAAA GCAAC CGCAG CACCT GCAAA GGCTG CATGG AAAAA ATTGA AAAAG
<< RMG014D_...  GATCC GAATT CAGAA TATGC GAAAA GCAAC CGCAG CACCT GCAAA GGCTG CATGG AAAAA ATTGA AAAAG
RMG009PL_20...        AT CAG-A TATGC GAAA- GCA-C CGCAG CACCT GCAAA GGCTG CATGG AAAAA ATTGA AAAAG
RMG007PC_20...         A TATGC GAAA- GCA-C CGCAG CACCT GCAAA GGCTG CATGG AAAAA ATTGA AAAAG
RMG004PL_20...                              GCAC CGCAG CACCT GCAAA GGCTG CATGG AAAAA ATTGA AAAAG
RMG008PL_20...                              GCAC CGCAG CACCT GCAAA GGCTG CATGG AAAAA ATTGA AAAAG
Construct_5...  GATCC GAATT CAGAA TATGC GAAAA GCAAC CGCAG CACCT GCAAA GGCTG CATGG AAAAA ATTGA AAAAG
                |1    |6    |11   |16   |21   |26   |31   |36   |41   |46   |51   |56   |61   |66
Contig1         GATCC GAATT CAGAA TATGC GAAAA GCAAC CGCAG CACCT GCAAA GGCTG CATGG AAAAA ATTGA AAAAG


<< RMG004D_...  GCCAG GTGCG CCTGA GCAAA AAAAT GGTGG ATCCG TATAC GCCCG CTGGA TGGCA TGATT GATCG CTGGT
<< RMG013D_...  GCCAG GTGCG CCTGA GCAAA AAAAT GGTGG ATCCG TATAC GCCCG CTGGA TGGCA TGATT GATCG CTGGT
<< RMG008D_...  GCCAG GTGCG CCTGA GCAAA AAAAT GGTGG ATCCG TATAC GCCCG CTGGA TGGCA TGATT GATCG CTGGT
<< RMG015D_...  GCCAG GTGCG CCTGA GCAAA AAAAT GGTGG ATCCG TATAC GCCCG CTGGA TGGCA TGATT GATCG CTGGT
<< RMG012D_...  GCCAG GTGCG CCTGA GCAAA AAAAT GGTGG ATCCG TATAC GCCCG CTGGA TGGCA TGATT GATCG CTGGT
<< RMG018D_...  GCCAG GTGCG CCTGA GCAAA AAAAT GGTGG ATCCG TATAC GCCCG CTGGA TGGCA TGATT GATCG CTGGT
<< RMG003D_...  GCCAG GTGCG CCTGA GCAAA AAAAT GGTGG ATCCG TATAC GCCCG CTGGA TGGCA TGATT GATCG CTGGT
<< RMG016D_...  GCCAG GTGCG CCTGA GCAAA AAAAT GGTGG ATCCG TATAC GCCCG CTGGA TGGCA TGATT GATCG CTGGT
<< RMG019D_...  GCCAG GTGCG CCTGA GCAAA AAAAT GGTGG ATCCG TATAC GCCCG CTGGA TGGCA TGATT GATCG CTGGT
<< RMG017D_...  GCCAG GTGCG CCTGA GCAAA AAAAT GGTGG ATCCG TATAC GCCCG CTGGA TGGCA TGATT GATCG CTGGT
<< RMG009D_...  GCCAG GTGCG CCTGA GCAAA AAAAT GGTGG ATCCG TATAC GCCCG CTGGA TGGCA TGATT GATCG CTGGT
<< RMG007D_...  GCCAG GTGCG CCTGA GCAAA AAAAT GGTGG ATCCG TATAC GCCCG CTGGA TGGCA TGATT GATCG CTGGT
<< RMG002D_...  GCCAG GTGCG CCTGA GCAAA AAAAT GGTGG ATCCG TATAC GCCCG CTGGA TGGCA TGATT GATCG CTGGT
<< RMG005D_...  GCCAG GTGCG CCTGA GCAAA AAAAT GGTGG ATCCG TATAC GCCCG CTGGA TGGCA TGATT GATCG CTGGT
<< RMG010D_...  GCCAG GTGCG CCTGA GCAAA AAAAT GGTGG ATCCG TATAC GCCCG CTGGA TGGCA TGATT GATCG CTGGT
<< RMG022D_...  GCCAG GTGCG CCTGA GCAAA AAAAT GGTGG ATCCG TATAC GCCCG CTGGA TGGCA TGATT GATCG CTGGT
<< RMG011D_...  GCCAG GTGCG CCTGA GCAAA AAAAT GGTGG ATCCG TATAC GCCCG CTGGA TGGCA TGATT GATCG CTGGT
<< RMG014D_...  GCCAG GTGCG CCTGA GCAAA AAAAT GGTGG ATCCG TATAC GCCCG CTGGA TGGCA TGATT GATCG CTGGT
RMG009PL_20...  GA-AG GTGCG CCTGA GCAAA AAAAT GGTGG ATCCG TATAC GCCCG CTGGA TGGCA TGATT GATCG CTGGT
RMG007PC_20...  GCCAG GTGCG CCTGA GCAAA AAAAT GGTGG ATCCG TATAC GCCCG CTGGA TGGCA TGATT GATCG CTGGT
RMG004PL_20...  GCCAG GTGCG CCTGA GCAAA AAAAT GGTGG ATCCG TATAC GCCCG CTGGA TGGCA TGATT GATCG CTGGT
RMG008PL_20...  GCCAG GTGCG CCTGA GCAAA AAAAT GGTGG ATCCG TATAC GCCCG CTGGA TGGCA TGATT GATCG CTGGT
Construct_5...  GCCAG GTGCG CCTGA GCAAA AAAAT GGTGG ATCCG NNKNN KCC-G NNKNN KGGCA TGATT GATCG CTGGT
                |71   |76   |81   |86   |91   |96   |101  |106  |111  |116  |121  |126  |131  |136
Contig1         GCCAG GTGCG CCTGA GCAAA AAAAT GGTGG ATCCG TATAC GCCCG CTGGA TGGCA TGATT GATCG CTGGT
```

38

```
<< RMG004D_...  ACCAT CCGGG CTG-- ----- AAAGG GGGGG AAGAA CTGGT GTGTC GCCCA GAATA TAGCG CGAGC CAGCT
<< RMG013D_...  ACCAT CCGGG CTG-- ----- AAAGG GGGGG AAGAA CTGGT GTGTC GCCCA GAATA TAGCG CGAGC CAGCT
<< RMG008D_...  ACCAT CCGGG CTG-- ----- AAAGG GGGGG AAGAA CTGGT GTGTC GCCCA GAATA TAGCG CGAGC CAGCT
<< RMG015D_...  ACCAT CCGGG CTG-- ----- AAAGG GGGGG AAGAA CTGGT GTGTC GCCCA GAATA TAGCG CGAGC CAGCT
<< RMG012D_...  ACCAT CCGGG CTG-- ----- AAAGG GGGGG AAGAA CTGGT GTGTC GCCCA GAATA TAGCG CGAGC CAGCT
<< RMG018D_...  ACCAT CCGGG CTG-- ----- -AAAG GGGGG AAGAA CTGGT GTGTC GCCCA GAATA TAGCG CGAGC CAGCT
<< RMG003D_...  ACCAT CCGGG CTG-- ----- AAAGG GGGGG AAGAA CTGGT GTGTC GCCCA GAATA TAGCG CGAGC CAGCT
<< RMG016D_...  ACCAT CCGGG CTG-- ----- AAAGG GGGGG AAGAA CTGGT GTGTC GCCCA GAATA TAGCG CGAGC CAGCT
<< RMG019D_...  ACCAT CCGGG CTG-- ----- AAAGG GGGGG AAGAA CTGGT GTGTC GCCCA GAATA TAGCG CGAGC CAGCT
<< RMG017D_...  ACCAT CCGGG CTG-- ----- AAAGG GGGGG AAGAA CTGGT GTGTC GCCCA GAATA TAGCG CGAGC CAGCT
<< RMG009D_...  ACCAT CCGGG CTG-- ----- AAAGG GGGGG AAGAA CTGGT GTGTC GCCCA GAATA TAGCG CGAGC CAGCT
<< RMG007D_...  ACCAT CCGGG CTG-- ----- AAAGG GGGGG AAGAA CTGGT GTGTC GCCCA GAATA TAGCG CGAGC CAGCT
<< RMG002D_...  ACCAT CCGGG CTG-- ----- AAAGG GGGGG AAGAA CTGGT GTGTC GCCCA GAATA TAGCG CGAGC CAGCT
<< RMG005D_...  ACCAT CCGGG CTG-- ----- AAAGG GGGGG AAGAA CTGGT GTGTC GCCCA GAATA TAGCG CGAGC CAGCT
<< RMG010D_...  ACCAT CCGGG CTG-- ----- AAAGG GGGGG AAGAA CTGGT GTGTC GCCCA GAATA TAGCG CGAGC CAGCT
<< RMG022D_...  ACCAT CCGGG CTG-- ----- AAAGG GGGGG AAGAA CTGGT GTGTC GCCCA GAATA TAGCG CGAGC CAGCT
<< RMG011D_...  ACCAT CCGGG CTG-- ----- AAAGG GGGGG AAGAA CTGGT GTGTC GCCCA GAATA TAGCG CGAGC CAGCT
<< RMG014D_...  ACCAT CCGGG CTG-- ----- AAAGG GGGGG AAGAA CTGGT GTGTC GCCCA GAATA TAGCG CGAGC CAGCT
RMG009PL_20...  ACCAT CCGGG CTG-- ----- AAAGG GGGGG AAGAA CTGGT GTGTC GCCCA GAATA TAGCG CGAGC CAGCT
RMG007PC_20...  ACCAT CCGGG CTG-- ----- AAAGG GGGGG AAGAA CTGGT GTGTC GCCCA GAATA TAGCG CGAGC CAGCT
RMG004PL_20...  ACCAT CCGGG CTG-- ----- AAAGG GGGGG AAGAA CTGGT GTGTC GCCCA GAATA TAGCG CGAGC CAGCT
RMG008PL_20...  ACCAT CCGGG CTG-- ----- AAAGG GGGGG AAGAA CTGGT GTGTC GCCCA GAATA TAGCG CGAGC CAGCT
Construct_5...  ACCAT CCGGG CTGCT TTGTG AAANN KNNKG AAGAA CTGNN KNNKC GCCCA GAATA TAGCG CGAGC CAGCT
               |141  |146  |151  |156  |161  |166  |171  |176  |181  |186  |191  |196  |201  |206
Contig1        ACCAT CCGGG CTG-- ----- AAAGG GGGGG AAGAA CTGGT GTGTC GCCCA GAATA TAGCG CGAGC CAGCT


<< RMG004D_...  GAAAG GCTTT AGCCT GCTGG CGACC GAAGA TAAAG AAGCG CTGAA AAAAC AGCTG CCGAA GCT-G CGGC
<< RMG013D_...  GAAAG GCTTT AGCCT GCTGG CGACC GAAGA TAAAG AAGCG CTGAA AAAAC AGCTG CCGAA GCT-G CGGC
<< RMG008D_...  GAAAG GCTTT AGCCT GCTGG CGACC GAAGA TAAAG AAGCG CTGAA AAAAC AGCTG CCGAA GCT-G CGGC
<< RMG015D_...  GAAAG GCTTT AGCCT GCTGG CGACC GAAGA TAAAG AAGCG CTGAA AAAAC AGCTG CCGAA GCT-G CGGC
<< RMG012D_...  GAAAG GCTTT AGCCT GCTGG CGACC GAAGA TAAAG AAGCG CTGAA AAAAC AGCTG CCGAA GCT-G CGGC
<< RMG018D_...  GAAAG GCTTT AGCCT GCTGG CGACC GAAGA TAAAG AAGCG CTGAA AAAAC AGCTG CCGAA GCT-G CGGC
<< RMG003D_...  GAAAG GCTTT AGCCT GCTGG CGACC GAAGA TAAAG AAGCG CTGAA AAAAC AGCTG CCGAA GCT-G CGGC
<< RMG016D_...  GAAAG GCTTT AGCCT GCTGG CGACC GAAGA TAAAG AAGCG CTGAA AAAAC AGCTG CCGAA GCTTG CGGC
<< RMG019D_...  GAAAG GCTTT AGCCT GCTGG CGACC GAAGA TAAAG AAGCG CTGAA AAAAC AGCTG CCGAA GCT-G CGGC
<< RMG017D_...  GAAAG GCTTT AGCCT GCTGG CGACC GAAGA TAAAG AAGCG CTGAA AAAAC AGCTG CCGAA GCT-G CGGC
<< RMG009D_...  GAAAG GCTTT AGCCT GCTGG CGACC GAAGA TAAAG AAGCG CTGAA AAAAC AGCTG CCGAA GCT-G CGGC
<< RMG007D_...  GAAAG GCTTT AGCCT GCTGG CGACC GAAGA TAAAG AAGCG CTGAA AAAAC AGCTG CCGAA GCT-G CGGC
<< RMG002D_...  GAAAG GCTTT AGCCT GCTGG CGACC GAAGA TAAAG AAGCG CTGAA AAAAC AGCTG CCGAA GCT-G CGGC
<< RMG005D_...  GAAAG GCTTT AGCCT GCTGG CGACC GAAGA TAAAG AAGCG CTGAA AAAAC AGCTG CCGAA GCT-G CGGC
<< RMG010D_...  GAAAG GCTTT AGCCT GCTGG CGACC GAAGA TAAAG AAGCG CTGAA AAAAC AGCTG CCGAA GCT-G CGGC
<< RMG022D_...  GAAAG GCTTT AGCCT GCTGG CGACC GAAGA TAAAG AAGCG CTGAA AAAAC AGCTG CCGAA GCT-G CGGC
<< RMG011D_...  GAAAG GCTTT AGCCT GCTGG CGACC GAAGA TAAAG AAGCG CTGAA AAAAC AGCTG CCGAA GCT-G CGGC
<< RMG014D_...  GAAAG GCTTT AGCCT GCTGG CGACC GAAGA TAAAG AAGCG CTGAA AAAAC AGCTG CC
RMG009PL_20...  GAAAG GCTTT AGCCT GCTGG CGACC GAAGA TAAAG AAGCG CTGAA AAAAC AGCTG CCGAA GCTTG CGGC
RMG007PC_20...  GAAAG GCTTT AGCCT GCTGG CGACC GAAGA TAAAG AAGCG CTGAA AAAAC AGCTG CCGAA GCTTG CGGC
RMG004PL_20...  GAAAG GCTTT AGCCT GCTGG CGACC GAAGA TAAAG AAGCG CTGAA AAAAC AGCTG CCGAA GCTTG CGGC
RMG008PL_20...  GAAAG GCTTT AGCCT GCTGG CGACC GAAGA TAAAG AAGCG CTGAA AAAAC AGCTG CCGAA GCTTG CGGC
Construct_5...  GAAAG GCTTT AGCCT GCTGG CGACC GAAGA TAAAG AAGCG CTGAA AAAAC AGCTG CCGAA GCTTG CGGC
               |211  |216  |221  |226  |231  |236  |241  |246  |251  |256  |261  |266  |271  |276
Contig1        GAAAG GCTTT AGCCT GCTGG CGACC GAAGA TAAAG AAGCG CTGAA AAAAC AGCTG CCGAA GCTTG CGGC
```

References

1. Janin, J., et al., *Macromolecular recognition in the Protein Data Bank.* Acta Crystallographica Section D: Biological Crystallography, 2006. **63**(1): p. 1-8.

2. Deremble, C. and R. Lavery, *Macromolecular recognition.* Current opinion in structural biology, 2005. **15**(2): p. 171-175.

3. Goodrich, J.A., G. Cutler, and R. Tjian, *Contacts in Context: Promoter Specificity and Macromolecular Interactions in Transcription.* Cell, 1996. **84**(6): p. 825-830.

4. Pommier, Y. and J. Cherfils, *Interfacial inhibition of macromolecular interactions: nature's paradigm for drug discovery.* Trends in Pharmacological Sciences, 2005. **26**(3): p. 138-145.

5. Sarai, A. and H. Kono, *PROTEIN-DNA RECOGNITION PATTERNS AND PREDICTIONS.* Annual Review of Biophysics and Biomolecular Structure, 2005. **34**(1): p. 379-398.

6. Pabo, C.O. and R.T. Sauer, *Protein-DNA Recognition.* Annual Review of Biochemistry, 1984. **53**(1): p. 293-321.

7. Petrucco, S., *Sensing DNA damage by PARP-like fingers.* Nucleic Acids Research, 2003. **31**(23): p. 6689-6699.

8. Saul, R. and B. Ames, *Background levels of DNA damage in the population.* Basic Life Sci, 1986. **38**: p. 529.

9. Knudson, A.G., *Hereditary cancer, oncogenes, and antioncogenes.* Cancer Res, 1985. **45**(4): p. 1437-1443.

10. Adimoolam, S. and J.M. Ford, *p53 and regulation of DNA damage recognition during nucleotide excision repair.* DNA repair, 2003. **2**(9): p. 947-954.

11. Luscombe, N.M., et al., *An overview of the structures of protein-DNA complexes.* Genome biology, 2000. **1**(1): p. reviews001.

12. Harrison, S.C., *A structural taxonomy of DNA-binding domains.* Nature, 1991. **353**(6346): p. 715-719.

13. Brennan, R. and B.W. Matthews, *The helix-turn-helix DNA binding motif.* J. biol. Chem, 1989. **264**(4): p. 1903-1906.

14. Garvie, C.W. and C. Wolberger, *Recognition of specific DNA sequences.* Molecular cell, 2001. **8**(5): p. 937-946.

15. Wolfe, S.A., L. Nekludova, and C.O. Pabo, *DNA RECOGNITION BY Cys2His2 ZINC FINGER PROTEINS.* Annual Review of Biophysics and Biomolecular Structure, 2000. **29**(1): p. 183-212.

16. Yang, W., *Structure and mechanism for DNA lesion recognition.* Cell Res, 2008. **18**(1): p. 184-197.

17. Petrucco, S. and R. Percudani, *Structural recognition of DNA by poly(ADP-ribose)polymerase-like zinc finger families.* FEBS Journal, 2008. **275**(5): p. 883-893.

18. Cotner-Gohara, E., et al., *Two DNA-binding and Nick Recognition Modules in Human DNA Ligase III.* Journal of Biological Chemistry, 2008. **283**(16): p. 10764-10772.

19. Taylor, R.M., C.J. Whitehouse, and K.W. Caldecott, *The DNA ligase III zinc finger stimulates binding to DNA secondary structure and promotes end joining.* Nucleic Acids Research, 2000. **28**(18): p. 3558-3563.

20. Plummer, E.R., *Inhibition of poly (ADP-ribose) polymerase in cancer.* Current Opinion in Pharmacology, 2006. **6**(4): p. 364.

21. Amé, J.-C., C. Spenlehauer, and G. de Murcia, *The PARP superfamily.* BioEssays, 2004. **26**(8): p. 882-893.

22. Langelier, M.-F., et al., *Crystal Structures of Poly(ADP-ribose) Polymerase-1 (PARP-1) Zinc Fingers Bound to DNA: STRUCTURAL AND FUNCTIONAL INSIGHTS INTO DNA-DEPENDENT PARP-1 ACTIVITY.* Journal of Biological Chemistry, 2011. **286**(12): p. 10690-10701.

23. Lilyestrom, W., et al., *Structural and biophysical studies of human PARP-1 in complex with damaged DNA.* Journal of Molecular Biology, 2010. **395**(5): p. 983-994.

24. Eustermann, S., et al., *The DNA-Binding Domain of Human PARP-1 Interacts with DNA Single-Strand Breaks as a Monomer through Its Second Zinc Finger.* Journal of Molecular Biology, 2011. **407**(1): p. 149-170.

25. Ali, A.A., et al., *The zinc-finger domains of PARP1 cooperate to recognize DNA strand breaks.* Nature structural & molecular biology, 2012. **19**(7): p. 685-692.

26. Jamieson, A.C., J.C. Miller, and C.O. Pabo, *Drug discovery with engineered zinc-finger proteins.* Nat Rev Drug Discov, 2003. **2**(5): p. 361-368.

27. Choo, Y. and A. Klug, *Selection of DNA binding sites for zinc fingers using rationally randomized DNA reveals coded interactions.* Proceedings of the National Academy of Sciences, 1994. **91**(23): p. 11168-11172.

28. Liu, Q., et al., *Design of polydactyl zinc-finger proteins for unique addressing within complex genomes.* Proceedings of the National Academy of Sciences, 1997. **94**(11): p. 5525-5530.

29. Beerli, R.R., B. Dreier, and C.F. Barbas, *Positive and negative regulation of endogenous genes by designed transcription factors.* Proceedings of the National Academy of Sciences, 2000. **97**(4): p. 1495-1500.

30. Carroll, D., *Progress and prospects: Zinc-finger nucleases as gene therapy agents.* Gene Ther, 2008. **15**(22): p. 1463-1468.

31. Urnov, F.D., et al., *Highly efficient endogenous human gene correction using designed zinc-finger nucleases.* Nature, 2005. **435**(7042): p. 646-651.

32. Smith, G.P. and V.A. Petrenko, *Phage display.* Chemical reviews, 1997. **97**(2): p. 391-410.

33. Brissette, R., J.K. Prendergast, and N.I. Goldstein, *Identification of cancer targets and therapeutics using phage display.* Current Opinion in Drug Discovery and Development, 2006. **9**(3): p. 363.

34. Rodi, D.J. and L. Makowski, *Phage-display technology – finding a needle in a vast molecular haystack.* Current Opinion in Biotechnology, 1999. **10**(1): p. 87-93.

35.   Levitan, B., *Stochastic modeling and optimization of phage display.* Journal of Molecular Biology, 1998. **277**(4): p. 893-916.

36.   Zwick, M.B., J. Shen, and J.K. Scott, *Phage-displayed peptide libraries.* Current Opinion in Biotechnology, 1998. **9**(4): p. 427-436.

37.   Jamieson, A.C., S.-H. Kim, and J.A. Wells, *In vitro selection of zinc fingers with altered DNA-binding specificity.* Biochemistry, 1994. **33**(19): p. 5689-5695.

38.   Greisman, H.A. and C.O. Pabo, *A General Strategy for Selecting High-Affinity Zinc Finger Proteins for Diverse DNA Target Sites.* Science, 1997. **275**(5300): p. 657-661.

39.   Wolfe, S.A., et al., *Analysis of zinc fingers optimized via phage display: evaluating the utility of a recognition code.* Journal of Molecular Biology, 1999. **285**(5): p. 1917-1934.

40.   Schuler, B. and L.K. Pannell, *Specific Labeling of Polypeptides at Amino-Terminal Cysteine Residues Using Cy5-benzyl Thioester.* Bioconjugate Chemistry, 2002. **13**(5): p. 1039-1043.

41.   Farmer, H., et al., *Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy.* Nature, 2005. **434**(7035): p. 917-921.