DNA LIGASE-CATALYZED SCAFFOLDING OF PEPTIDE FRAGMENTS ON NUCLEIC ACID POLYMERS

by

CHUN GUO

(Under the Direction of RYAN HILI)

ABSTRACT

By taking advantage of ssDNA self-folding into stable secondary structures, ssDNA can be applied as a scaffold for presenting peptide residues to mimic the interface of PPIs. We have developed a method for the T4 DNA ligase-catalyzed polymerization (LOOPER) of 5′-phosphorylated pentanucleotides containing peptide fragments. This polymerization proceeds sequence-specifically to generate DNA-scaffolded peptides in excellent yields. It has been shown to tolerate peptides ranging from two to eight amino acids in length with a wide variety of functionalities. We validated the capabilities of this system in a mock selection for the enrichment of a His-tagged DNA-scaffolded peptide phenotype from a library, where almost 190-fold enrichment after one round of selection was observed. A high-throughput duplex DNA sequencing method was developed to facilitate the determination of fidelity for various codon sets and library sizes. With this process, we identified several codon sets that enable the efficient and sequence-specific incorporation of peptide fragments along a ssDNA template at fidelities up to 99%. These

findings marked a significant advance in generating evolvable biomimetic polymers which should find ready application for the in vitro selection of molecular recognition.

INDEX WORDS: DNA-scaffold, protein-protein interactions, peptide fragments, SELEX, fidelity, ligase-catalyzed polymerization, PCR, bio-conjugation

DNA LIGASE-CATALYZED SCAFFOLDING OF PEPTIDE FRAGMENTS ON NUCLEIC ACID POLYMERS

by

CHUN GUO

BS, Shandong University, 2013

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2018

DNA LIGASE-CATALYZED SCAFFOLDING OF PEPTIDE FRAGMENTS ON NUCLEIC ACID POLYMERS

by

CHUN GUO

Major Professor: Ryan Hili

Committee: Vladimir Popik

Geert-Jan Boons

Electronic Version Approved:

Suzanne Barbour

Dean of the Graduate School

The University of Georgia

December 2018

DEDICATION

I would like to dedicate this work to my parents Liangguo Guo, Xuehua Yong and my beloved one Xin Wen for their support and love.

ACKNOWLEDGEMENTS

I would like to first thank my PI Prof. Ryan Hili for allowing me to join in his group and work on these interesting projects. Even being relatively new to synthetic chemical biology, I get a lot of hands on experience from Ryan. With the broad knowledge he has, Ryan also inspires me a lot for independent thinking. I really appreciate the guidance and help that Ryan has provided as my primary instructor.

In addition, I would like to thank my committee members Prof. Valdimir Popik and Prof. Geert-jan Boons in viewing my work and providing valuable suggestions to my projects.

I also appreciate my lab mates for the help and encouragement. Being the first in lab, Dr Delaney Hook always keeps the lab in a well-organized condition and takes the responsibility whenever anything goes wrong. Working in the similar field, Yi and Dehui always provide valuable suggestions regarding the problem I have encountered. Able to work with these people is my great pleasure and honor.

Finally, I would like to express my thanks to my friends Weiyu Li and Weizhong Zhang. The good times that we had playing together, discussing together and enjoying our lives in UGA will always be memorable for me. Special thank for Christopher Harrington for helping me editing the manuscript.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF SCHEMES

Page

CHAPTER 1

Introduction and literature review

1.1 Protein-protein interactions (PPIs) inhibitors

Proteins are biomacromolecules consisting of chains of amino acids residues. Interactions between proteins play a central role in a vast number of biological processes, including dysregulation in disease. It has brought the interest for researchers to identify inhibitors for protein-protein interactions (PPIs).[1,2,3] Protein-protein interfaces usually present certain features. Large surfaces with shallow topologies spanning from 1,500—3,000 Å[2,4] are often observed. Interactions usually occur at binding pocket generated from proper folding.[5] Amino acids (Trp, Tyr and Arg) are typically found at the interface which are correlated with hydrophilic and hydrophobic interactions.[6]

Guided by the features above, researchers have identified various inhibitors that can interrupt the PPIs. Based on the molecular weight, they can be divided into three types: small drug-like molecules (M.W.<=500 Da), protein epitope mimetics(0.5kDa-5kDa) and recombinant macromolecule (M.W. 10-200 kDa).[7] The first two will be mainly discussed here.

Small drug-like molecules

In the early stages, PPIs were not recognized as "druggable" by small molecules. As the contacting surface for proteins are a lot larger than that of small molecule ligand.[8] In addition, unlike G protein-coupled receptors (GPCRs) which have native small molecule ligands,[9] most PPIs don't have a native small molecule ligand to start the chemical

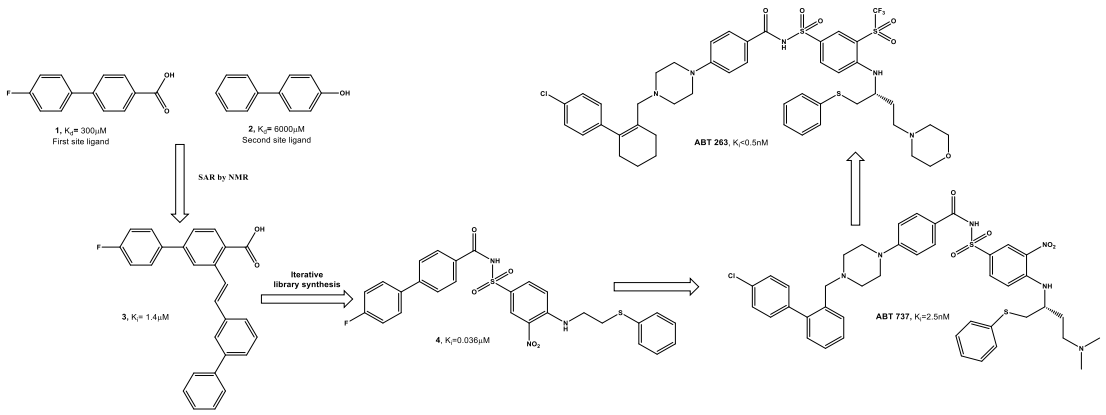
discovery process. Techniques, such as high throughput screening (HTS) of small molecule libraries, presented some difficulties in generating validated hits.

However, the discovery of several natural products such as rapamycin[10] and cyclosporine[11] revealed that small molecules could be effective inhibitors for PPIs. In addition, the identification of key roles for protein hot spots at PPIs rather than the complete interface[12] further enhanced the possibility.

During the meantime, the application of fragment-based-drug-discovery (FBDD) brought the hypothesis into reality. FBDD involves the assembly of weakly binding small molecules (M.W.<300 Da with mM binding) to achieve a potential drug candidate with nM binding.[13] FBDD typically consists of three key steps: (*i*) Preparing fragment libraries containing small molecule members with less than 300 Da molecular weight, less than three hydrogen bond donors and less than three hydrogen bond acceptors; (*ii*) Screening of the fragment libraries towards a biological target for possible non-covalent binding; (*iii*) Elaborating the fragment hits to achieve high bio-affinity results as to prepare potential therapeutic candidates.

The discovery of ABT-263(anti-apoptotic inhibitor of Bcl-2 family) was a typical example of apply of this strategy (**Scheme 1.1**).[14] Compound **1** and compound **2** were first identified through screening a 10000 membered fragment library and a 3500 membered fragment library respectively towards Bcl-xL.[15] Guided by SAR and NMR, these two fragments were assembled and redesigned to generate compound **3** with improved K_i of 1.4 μM. Further iterative library screening and synthesis led to compound **4** with K_i of 0.036 μM.[15] The early candidate **ABT 737** was identified through substituent optimization aiming at decreasing undesired binding to fetal bovine serine (FBS) and

increasing binding to other Bcl family members.[16] The final candidate **ABT 263** was generated through additional optimization based on medicinal chemistry aiming at improving oral bioavailability.[17]



Scheme 1.1: Small molecule inhibitor identified for Bcl-xL through FBDD.

FBDD showed better control over the size, complexity and physical properties of the small molecule. In addition, with the help of highly sensitive biophysical technique such as nuclear magnetic resonance (NMR),[15,18] mass spectroscopy (MS) [19,20,21] and X-ray crystallography[22,23], FBDD enabled the detection of weakly fragment binding. Finally, FBDD worked better in principle. As multiple protein hot spots were presented on the interface, it would be easier to identify fragments targeting individual hot spot than identifying small molecules that target multiple hot spots at the same time.

Obstacles are still present within FBDD. As structural information is required when determining protein-ligand binding, proteins with unknown structures would be difficult to screen. Also, the high cost and special expertise required in running the detection techniques impeded this application.[24]

Protein epitope mimetics

With a chemical diversity of only 20 amino acids, proteins still present various biological functions. Not only the composition but also the conformation of how they fold matters. Protein epitope mimetics stand for synthetic molecules that possess constrained structure that could mimic the three dimensional structure of a folded protein epitope which is recognized by its receptor.[25] In the early stages, success was achieved through preparing of β-turn peptide mimetics.[26] The β-turn was achieved through inserting a covalent linkage between amino acids at the strategic locations (**Figure 1.1**).[27]

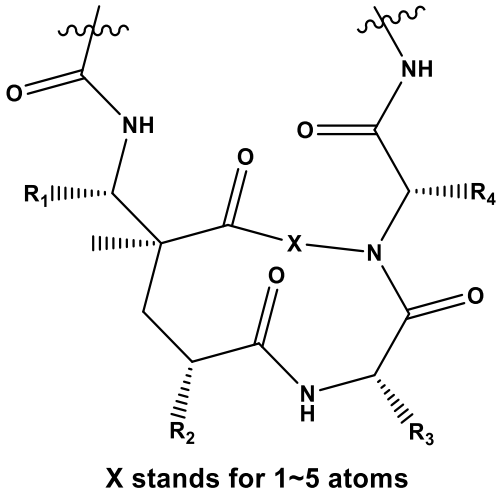


Figure 1.1: Early stage β-turn mimetics.

During the recent decades, other epitopes of interest have developed. Among which included α-helix,[28] β-sheet,[29] β hairpin[30] and other discontinued epitope[31] regions. Taking α-helix as an example, it has around 3.6 residues per turn with a vertical distance of 0.54nm between two consecutive residues. Residues in α-helices typically adopt backbone (φ, ψ) dihedral angles around (−60°, −45°). This helical conformations were found in over 30% of the protein secondary structure,[32] including p53/HDM2 interactions,[33] Bcl-xL/BIM interactions[34] and Mcl-1/ NOXA B interactions[35]. It makes them to be desired epitope for designing mimetics.

Based on the structural factors, α-helix typically adopted three phases and residues on position i, i+4 and i+7 are proximately on the same phase.[28] Guided by that, a set of synthetic templates have be generated (**Figure 1.2**).

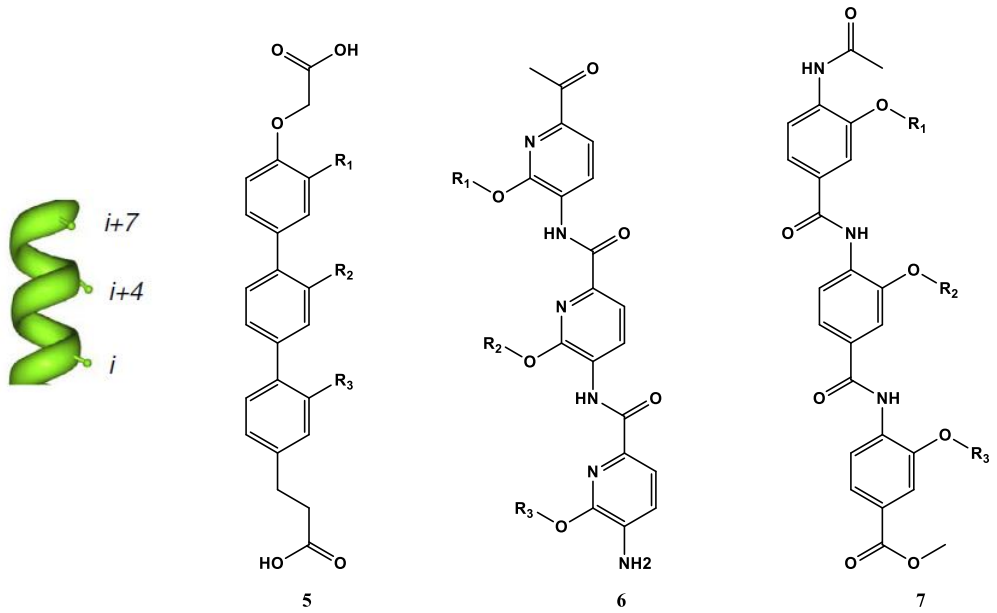


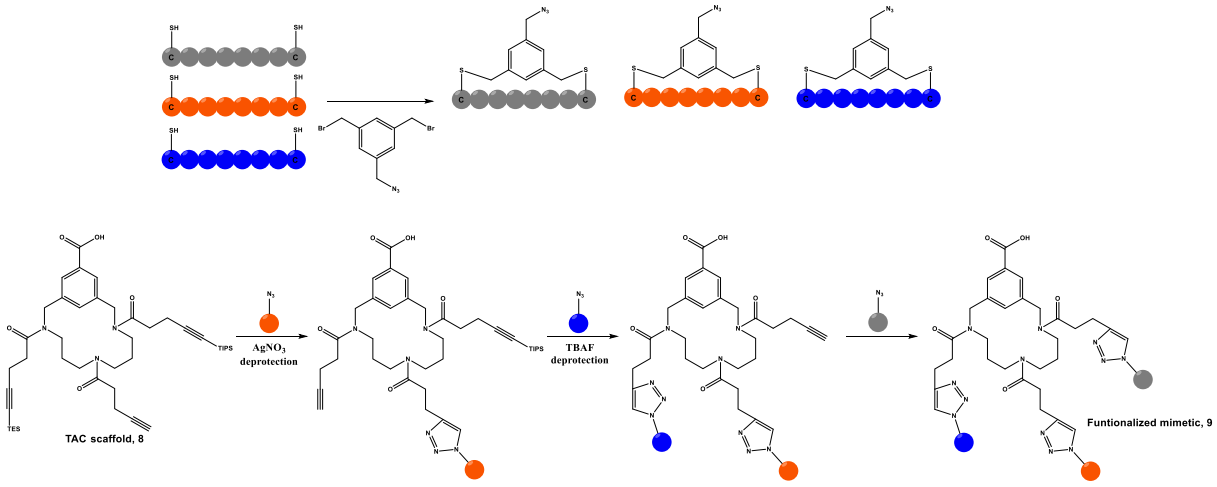
Figure 1.2: Organic molecule template for α-helix mimetics.

Helicity is structurally presented in organic compounds such as terphenyl and arylamide.[36] That allows them to efficiently mimic α-helix confirmation. Terphenyl compound **5** was first applied as the template for identifying potential PPI inhibitor for Bcl-xL. Val, Leu, Ile, Tyr and Phe residues, which are proved to be critical in Bcl-xL and Bak recognition[37] were chosen to be presented on the functionalization site. With that, one α-helix mimetic targeting Bcl-xL with a K_d of 114nm was identified.[38] The hydrogen bond between amide hydrogens towards pyridyl nitrogens and alkoxy groups enables the ability for compound **6** to mimic this helical structure. Overlay of polyaniline α-helix with compound **6** present a close resemblance with each other.[39] A computational study revealed that trisbenzamide compound **7** resembles α-helix confirmation even better as

the orientation of functional residues fits well with the residue on i, i+4 and i+7 position on an ideal α-helix conformation.[40]

While not all protein epitopes are composed of continued regions. A crystal structure of the CD4 receptor to gp120 have revealed 3 isolated epitopes.[41] In that way, other synthetic scaffolds were prepared to mimic this discontinued epitope.

Orthogonally protected TAC scaffold (compound **8**) was synthesized through the following protocol[42]. Peptide epitopes generated from the interface were cyclized through conjugating two terminal cysteine with benzylic dibromide (**Scheme 1.2**). Since the larger silyl ether protecting groups have higher resistance to hydrolysis, triethylsilyl ether (TES) and triisopropylsilyl ether (TIPS) could then be deprotected and sequentially coupled to a different cyclic peptide epitope to generate the mimetics (**Scheme 1.2**). Fully functionalized TAC scaffold (compound **9**) with cyclic peptide epitope enables a strong interaction to CD4 with a IC_{50} value of 41μM.[41]



Scheme 1.2: Preparation of cyclic peptide epitope and final functionalized TAC scaffold.

With the precise local environment presented, protein epitope mimetics enabled an effective resemblance of the protein interface. However, lacking of the ability to

synthesize through a combinatorial approach[43] limited it application in high-throughput screening[44], as each library member needs to be synthesized separately. One approach that could present the multiplexity of these functional groups and assemble them in a combinatory synthetic pathway would be of great potential for the discovery of new types of inhibitor for PPIs.

1.2 Nucleic acids guided display

The mechanism which governs the architecture of DNAs is well understood. In that, applying DNA as templates or scaffolds enables the generation of well-controlled secondary structures.[45,46] In addition, with the application of DNA sequencing and PCR amplification, DNA scaffold could be prepared and analyzed efficiently.[47] Among all the secondary structures, DNA's double helix is the most frequently observed DNA secondary structure. Following "Watson-Crick base pairing" rules, two single stranded DNAs hybridize into one double stranded DNA in helical form.

Double stranded DNA is found to be in three main conformations, of which B-DNAs are predominately presented, especially in cellular environment.[48] The double helixes of B-DNA are right handed. Each nucleobase has a distance of 3.4 Å to another nucleobase with a rotation degree of 34.3º. One complete turn requires around 10 bps and extended in 34 Å.[49] The double helixes of A-DNA are right handed as well which is similar to B-DNA while the helical structures are more compact compared to B-DNA. As the base pairs are more twisted and not perpendicular to each other. In that the rise per helix for A-DNA is 28.6 Å which is shorter than that of B-DNA.[50] The dehydration of DNA drives it to become A-form as to protect DNA under that conditions.[51] Z-DNA is in left-handed

helical form with a zig-zag backbone conformation. Higher G-C contents favor that form.[52] Higher order of helical structure are also presented but will not be discussed here.[53,54]

DNA scaffolded small molecules

DNA scaffolds offer spatial control over the orientation of small molecule fragments in targeting protein of interest (**Figure 1.3**). Each DNA strand is covalently conjugated to a small molecule fragment and contains a tag for that molecule. After affinity pull down, the desired combination of fragments will be identified from sequencing the DNA template which encodes for that molecule. Then covalent assembly of those fragment winners will be performed and tested. This procedure was widely applied in facilitating fragment base drug discovery.

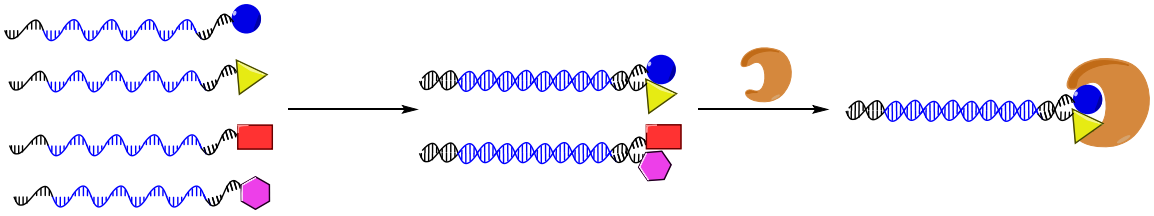


Figure 1.3: DNA scaffolds in identifying fragments assembles targeting protein of interest.

For the early stage discovery, Hamilton et al. assembled two distinct protein binding agents through covalent labeling to the end of each DNA strand.[55] In this characterization study, 2-iminobiotin which is a reversible binder of streptavidin was mixed into the library of small molecules. After coupling these small molecules to the DNA templates, annealing and affinity pull down with streptavidin bead were performed. Only the DNA duplex containing two 2-iminobiotin was detected as the binder towards streptavidin.

This proved the importance of the proximity for bivalent or multi-valent binding that DNA scaffold could provide.

Following this strategy, multiple bivalent ligands were identified with nanomolar range binding towards protein targets including streptavidin[56] (K_d 1.9nM) and trypsin[57] (IC_{50} 3 nM).

DNA scaffolded peptides

Even being rigid, DNA double helix still presents certain degree of multiplexity. The proximity and orientation of separate ligands on the double helix could be tuned by alternating the position of ligands (**Figure 1.4**).

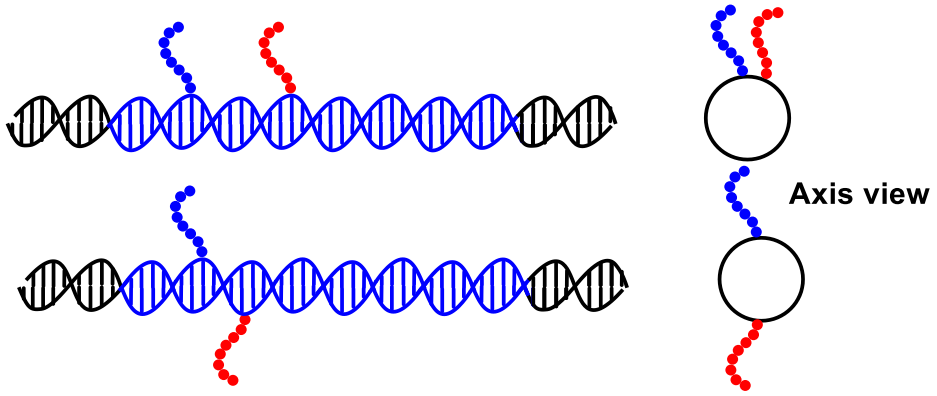


Figure 1.4: DNA display of ligand with various proximity and orientation.

Chaput et al. developed an approach for generating antibody mimetics-"synbody"[58] taking the advantage of multivalent interaction which is presented in native protein-protein interactions. A fragment-based approach was performed to first identify couple 12-mer peptides from a ~4000 membered peptide library that interact with the protein target (Gal80 protein) with micromolar affinity. Then a binding epitope mapping was performed against the peptides and protein target. With that approach, two most potent ligands that interact with the protein on different epitope were identified. They are BP1

with the K_d of 3.3μM and AD1 with the K_d of 3.6 μM. Following that those two ligands were then conjugated to variable positions on each strand of the dsDNA to access the divalent interaction. By alternating the position of one ligand while keeping the other fixed, Chaput et al. enabled the screening of scaffolds for displaying peptides and found a synbody that interacts with the protein target at K_d of 5.6 nM (~1000 fold increase in binding affinity). Removing any of the ligands rendered the K_d back to micromolar range. Later Chaput et al. showed that, applying mRNA display of peptide ligands towards target of interest instead of peptide array enables the identification of potential ligands with mid nanomolar range binding.[59] As increasing the library size enables the identification of better ligands.

The display of peptide ligands on dsDNA enables the generation of synbody with strong binding affinity. In addition, the generation of peptide ligands could either be achieved through peptide array or mRNA display which are both well validated approaches.

DNA double helical structures are relatively rigid and stable, inserting a nick site into one strand increased the flexibility through introducing the state of "open" and "close" (**Figure 1.5**).[60]

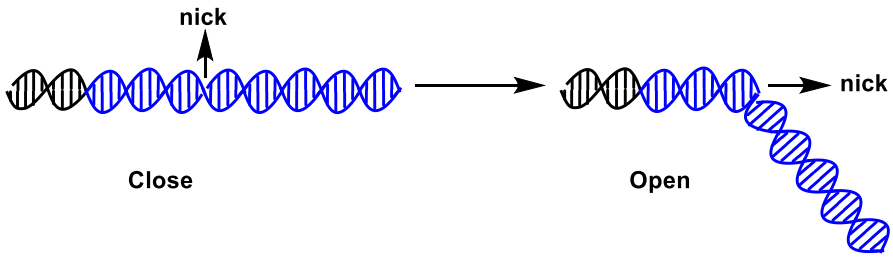


Figure 1.5: "Open" and "Close" state for nick containing dsDNA.

The proximity and orientation of functional residues could alter the biological functions as well. With that concept, DNA scaffold may function as molecular ruler for controlling

the proximity and orientation of multiple ligands as to achieve the better binding (**Figure 1.6**).

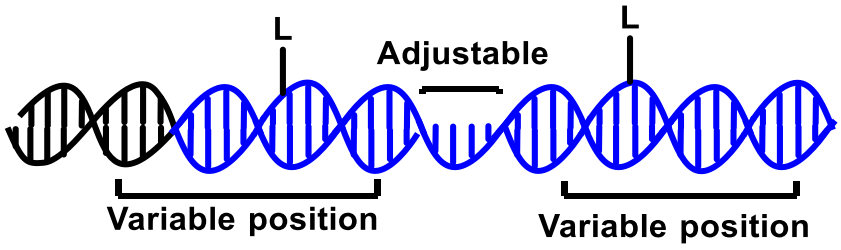


Figure 1.6: DNA controlled multivalent ligand presentation.

In this approach, ligands of interest will be covalently linked to the single stranded nucleotides oligomers. Following the base-pairing rule, the oligomers which contain the ligand of interest will be annealed to their corresponding position, certain length of ssDNA nick site will be introduced between the two double stranded DNA segments. By adjusting the position of ligands in the oligomer and the length for ssDNA nick, various assembly of ligand conformation could be achieved.

In the initial research, Seitz et al.[61] compared dsDNA scaffold and dsDNA with ssDNA nick scaffold in displaying two peptide ligands toward Syk kinase. They found that not only the distance between the ligands altered the binding, the flexibility of spacer in between also had an impact. The rigid helical twist presented in dsDNA scaffold was problematic for efficient display of ligands in protein binding, while having uncoupled ssDNA spacer in between allowed for the relief of the torsion constraint.

In a more recent study, Seitz et al. applied this DNA scaffolded spatial screening into distinguishing two proteins which share similar binding to a consensus sequence motif.[62] The tandem-SH2 domains (tSH2) of the Syk and ZAP-70 kinases both bind to a tyrosine phosphorylated motif pYXXI/L$(X)_{6-8}$pYXXI/ L with nanomolar K_d.[63] The underlined

parts are proved to be critical for interaction. Based on that, this motif was dissected into two parts containing pYXXLG. These two oligo peptide fragments were then coupled to the single strand nucleotides which gets annealed to the templates. The flexibility were controlled through the coupling orientation of peptides to nucleotides (from C'-terminal or N'-terminal), the relative orientation of the two peptides (C'-term to C'-term; C'-term to N'-term; N'-term to C'-term; N'-term to N'-term), and the relative distance between the two peptides(the spacer of dsDNA segments and ssDNA segments). The screening revealed that, compared to ZAP-70, Syk tSH2 has a rather broad substrate scope. In other words, the proximity and orientation of the two ligands does not have a significant impact on efficient binding. While ZAP-70 tSH2 requires proximal arrangement of the two ligands, only presenting the peptides in a relatively close proximity resulted in efficient binding. As other combination of the peptide ligands bound to Syk tSH2 tightly but failed to bind to ZAP-70 tSH2.

However, even with the presence of a nick site, the conformation varieties of dsDNA are not comparable to ssDNA. ssDNAs can generate more complex secondary structures including stems, loops and bulges. These spherical structures[64] result in close assembly of multiple ligands while dsDNAs remain to be rod-like in most cases. For multivalent interactions exceeding two, the proximity of ligands is simply controlled by the contents of ssDNA segments. In contrast, the presence of dsDNA fractions prohibited the efficient binding towards targets.[65]

1.3 Native and modified nucleic acids

For a long time, ribonucleic acids (RNAs) and deoxyribonucleic acids (DNAs) were recognized mostly as genetic codes which guide the growth and development of living beings.[66] The ability for them to fold into various stable secondary structure, like ribozymes which enables catalyzing reaction[67] has drawn the attention for researchers in perusing its non-biological functions. It was until the identification of polymerase chain reaction (PCR)[68] which enables the regeneration of large quantity of nucleic acids (NAs) from diluted NAs libraries that the evolution on nucleic acid became possible. In 1990, Gold et al[68] first presented the idea of **S**ystematic **E**volution of **L**igands by **EX**ponential enrichment (SELEX) (**Figure 1.7**) in evolution of nucleic acids targeting their corresponding protein-T4 DNA polymerase. The initial nucleic acid library was first generated then subjected to target binding; unbounded library members were then washed away; bounded library members were eluted and then submit to library regeneration through PCR; after iterative rounds of the same process, nucleic acids which interact to the target in the desired way could be identified when comparing of sequencing results of the initial library members and final library members.

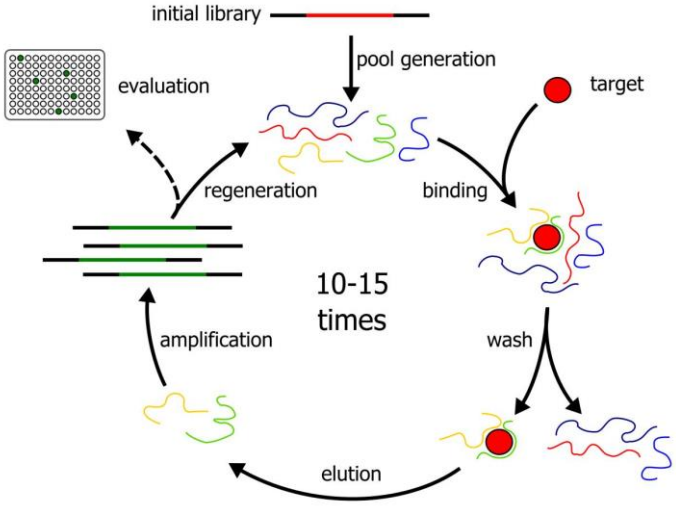


Figure 1.7: Systematic evolution of ligands by exponential enrichment.

With the similar approach Joyce et al[69] and Szostak et al[70] were able to identify the desired nucleic acids either cleaving specific DNA sequence[69] or targeting small molecule of interest[70] in the same year respectively. Also, that was the first time Ellington and Szostak introduced the word "aptamer" derived from the Latin word "aptus" as "to fit" to present single stranded nucleic acids (ssNAs) which are able to binding tightly and specifically to the target of interest.[70]

Traditionally, antibodies[71, 72]were applied in general for protein recognition and binding. Specifically, the tight binding to foreign invaders enables activation of immune systems as their key function.[73] By coupling to drug molecules, antibody-drug conjugates (ADC) enabled the targeted drug delivery which was applied in designing cancer therapeutics.[74, 75] However, they suffer from a number of shortcomings, including slow production time, high cost, limited target availability, poor stability and variable production quality.[76] Despite over 500,000 commercially available antibodies, researchers continue to express concerns over accessibility to high-quality antibodies for biomedical research.[77, 78, 79] These issues have stimulated the development of alternative technologies to generate high-affinity reagents for proteins that rival the performance of traditional antibodies.[80] Compared to antibodies, applying nucleic acids as an affinity reagents presents several advantages: (*i*) their generation through multiple rounds in vitro selection enables tuning of binding and specificity properties;[81] (*ii*) their active structure can be reversibly generated by thermal denaturation and cooling;[82] (*iii*) they exhibit excellent chemical stability and shelf-life[83] and (*iv*) their chemical synthesis is predictable and scalable.[84] Toole et al identified couple single-stranded DNAs (ssDNA) targeting human thrombin from a 10^{13} 96-mer library with K_d ranging from 25nM to 200nM.[85] Native aptamers

targeting HIV reverse transcriptase,[86] fibroblast growth factor 2,[87] vascular endothelial growth factor,[88] even small molecules as D-adenosine[89] and caffeine[90] have been evolved through the SELEX process.

When compared to antibodies, native aptamers still present a deficiency-lack of functionality. Proteins contain 20 amino acids while nucleic acids only have four (taking DNA as an example) (**Figure 1.8**). The limited functionalities potentially inhibit its application to various protein targets, as hydrogen bonds, the hydrophobic effect and interactions between positively and negatively charged residues are widely presented on the PPIs interface.[91] It would be difficult to target those which are negatively charged and hydrophobic in nature.[92] Identifying an approach to generate modified nucleic acids is demanded in pushing the application of aptamers to the next level.

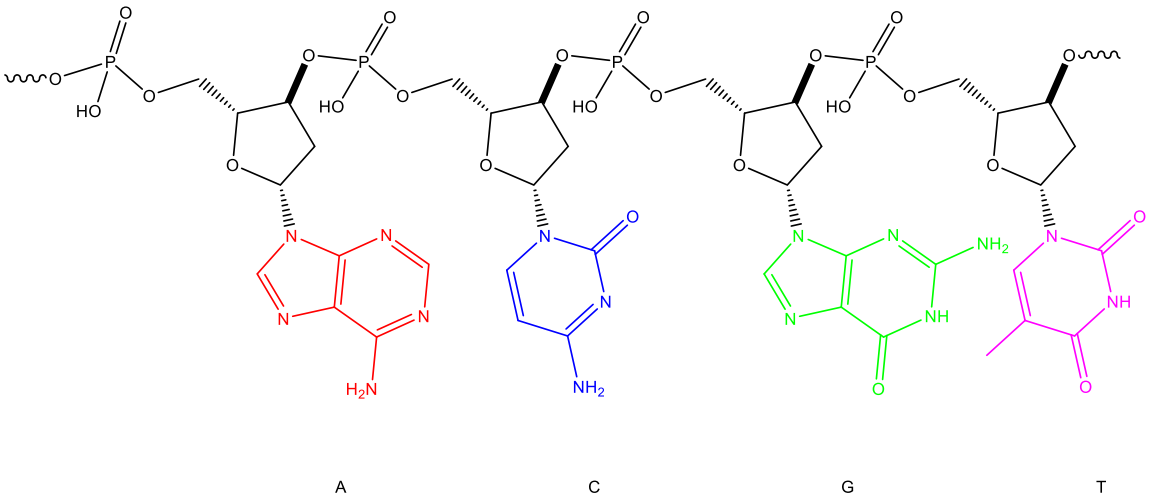
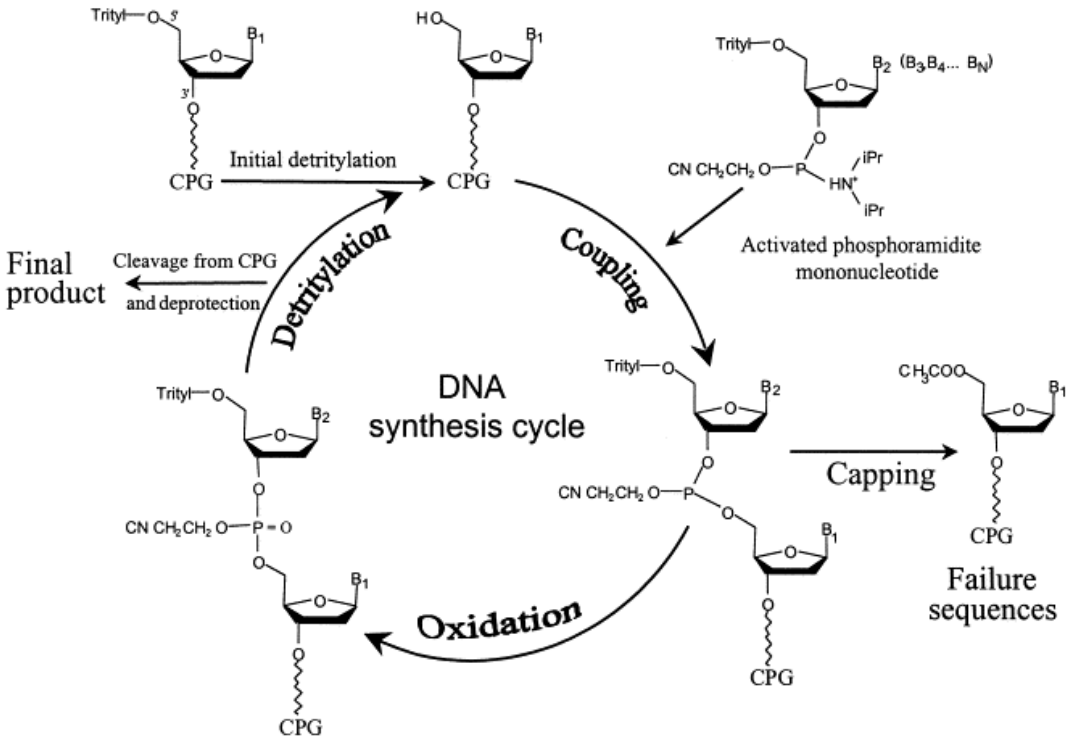


Figure 1.8: Limited functionality of native DNA.

Functionalized nucleic acid aptamer through chemical synthesis

The establishment of solid phase DNA synthesis allows the synthesis of nucleic acid in preparative scale. Solid phase DNA synthesis is performed through following this procedure (**Scheme 1.3**) (*i*) Removing the trityl protecting group to free 5'-hydroxyl with

3% trichloro acetic acid; (*ii*) Activating the phosphorous atom of the phosphoramidite with tetrazole, then the free 5'-hydroxyl group would attack it and get coupled to the solid support, (*iii*) Capping the uncoupled 5'-hydroxyl group with the mixture of acetic anhydride and *N*-methylimidazole. In that way, the unreacted the 5'-hydroxyl group is quenched and won't be reactive to the next phosphoramidite; (*iv*) Oxidizing the phosphite triester to phosphate through I_2 to generate stable phosphate diester bond. Through iterative rounds of cycles, oligonucleotide with the desired sequence is generated. After cleaving the oligomers off the beads and purification, synthetically generated oligonucleotides are prepared.



Scheme 1.3: Solid phase DNA synthesis procedure.[93]

Taking that process into consideration, introducing modified phosphoramidites should enable preparation of functionalized nucleic acids through solid phase oligonucleotide synthesis. Commercially available modified phosphoramidites contain a reactive head

including an amino residue, a carboxylate residue and an alkyne residue which are used for efficient labeling after synthesis. Customized synthesis of specifically functionalized phoshphoramidite could be time consuming and expensive at the same time.[94] Also, the positions of modification for the nucleic acid generated from solid phase oligonucleotide synthesis are fixed. A library of that already loses a significant amount of variability which is disfavored to be applied for in vitro evolution. Instead, that approach would be more suitable in preparation for large quantity of post-SELEX modified aptamer for binding kinetics examination.[95,96]

Functionalized nucleic acid aptamer through modified nucleotide triphosphate

RNAs and DNAs are generated through RNA polymerase[97] or DNA polymerase[98] based on a parent DNA template using ribonucleotide triphosphate (NTP) or deoxyribonucleotide triphosphate (dNTP) substrates respectively. When applying modified substrates, it is possible to prepare modified nucleic acid enzymatically. Majority of the native polymerase present a high resolution towards different nucleoside bases. For instance, taq polymerase makes one mistake in copying every 10000 bps, the error rate was 0.01%.[99] That prevents the mis incorporation of the nucleotide as to keep the accuracy when copying from a template. However, the high-resolution on the other hand may also prohibit the recognition of modified nucleotide triphosphate substrate. Introducing functionality onto the nucleoside base enables a broad range of chemical diversity. The position for introducing the modification are critical, of which the most important is to preserve the native Waston-Crick geometry.[100] It has been shown that NTP with modifications on 7' position (C7) of deazapurines and 5' position (C5) of pyrimidine (**Figure 1.9**) remains its activity of substrate towards polymerase.[101] In

addition, crystal structures of C7/C5 modified NTP with polymerase, template and primer indicated that these modified NTPs remain in the catalytic center for the enzyme.[102]

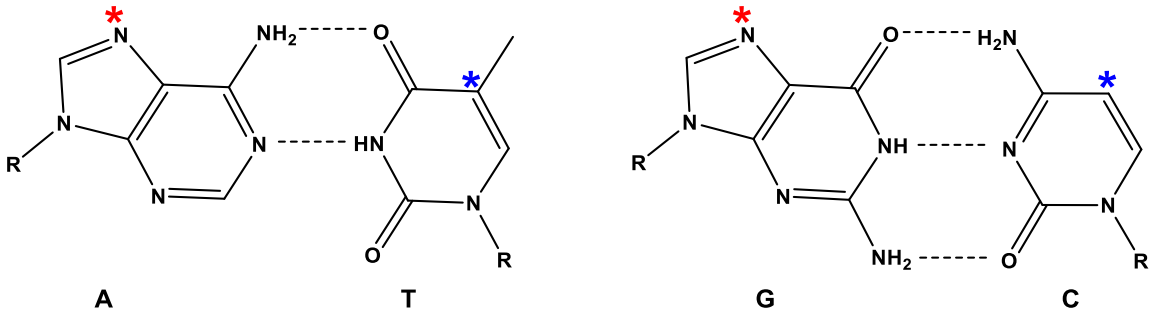


Figure 1.9: Modification positions of NTPs (modified position was marked as * and *).

Toole et al applied 5-(1-pentynyl)-2'-deoxyuridine(dU*TP) in SELEX against thrombin.[103] With that, they managed to isolate a second class of thrombin aptamer pool with different sequence composition compared with initial aptamer derived using native dNTPs.[104] The target binding relied on the presence of modified uridine as removing that resulted loss in binding. This initiated the approach for evolving aptamers with various functionalities.

Intended to expand chemical diversity for aptamers, researchers in SomaLogic,Inc., introduced amino-acid side chain like modifications through modified dU*TP to create Slow-Off rate Modified Aptamers (SOMAmers).[105] An array of modified dN*TPs was prepared for the generation of SOMAmers (**Figure 1.10**). Most of the modifications contain aromatic residues like benzyl and naphthyl groups. The hydrophobic interaction helps increase the interaction between SOMAmers and protein target, as hydrophobic amino acid residues like Trp and Tyr were recognized as hot-spot in protein-protein interface.[91] In addition, the modification were introduced through palladium catalyzed carbonylative coupling with primary amines and CO[106] which would leave an amide

linkage. This amide linkage allows the formation of additional hydrogen bonds which help stabilizing aptamer secondary structure.

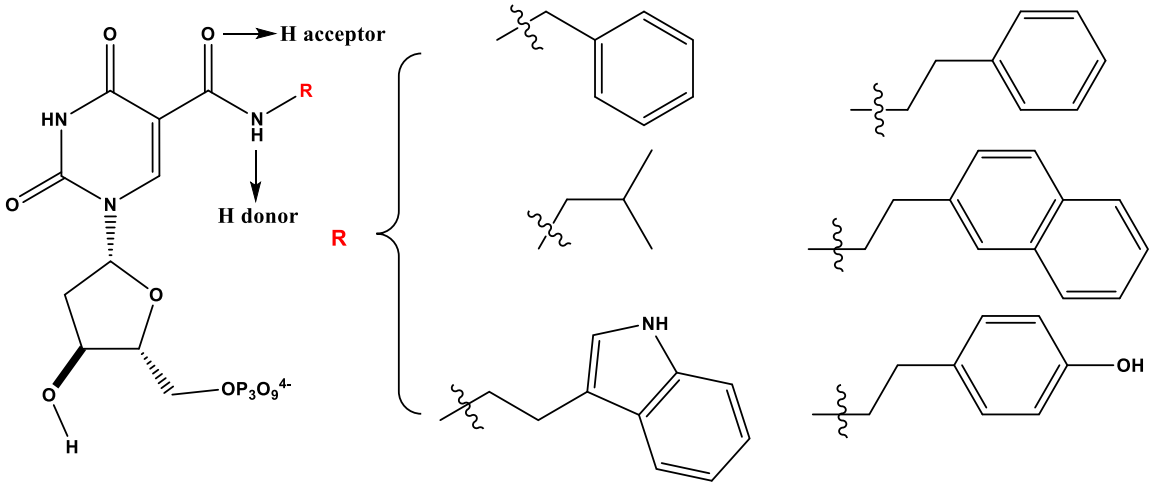


Figure 1.10: Some examples of modified dU*TP applied for SOMAmers.

With the modification, SOMAmers were able to specifically recognize protein biomarkers including IL-8, tPA, resistin, MIP-4, MMP-7, MMP-9, RANTES, MCP-1, and Lipocalin 2 respectively.[105] Recently Gawande et al presented a new approach which applied functionalized dC*NP and dU*TP at the same time.[107] In that way, two different modifications will be presented on the nucleic acids. Using two different functional groups instead of one, a SOMAmer targeting PCSK9 with K_d<1nM was identified. Aside from high affinity, it has been shown that the presence of second modification didn't interrupt with specific recognition. As the SOMAmer didn't show any binding towards PCSK1, PCSK2, furin, PCSK4, or PCSK7 (13.5–16.2% identity to PCSK9) even at 100nM. This further proved the importance of expanding chemical diversity in aptamer application.

The size of modification on dN*TP might also be an important factor in order to be efficiently recognized by polymerase, however, the limitation of the molecule size does

not seem straightforward. Despite the capability of incorporating dN*TP with large molecules like fluorescent dyes,[101] Marx et al. showed that a 23nt oligomer modified dT*TP could be recognized and incorporated onto a DNA template consecutively up to seven by *Therminator* DNA polymerase. Even though the modified dNTPs are 40 times larger than the native substrate.[108] Years later, horseradish peroxidase (HRP) modified dT*TP which is 100 times larger than the native substrate got incorporated by *KlenTaq* polymerase[109]. Recently, even antibody modified dC*TP which is larger than the polymerase still got incorporated by KOD DNA polymerase.[110] Even with such promise, applying extremely large modified dN*TP in SELEX would still be problematic. Firstly, it has been found that the polymerization efficiency of modifier depends on the neighboring nucleotide. That would decrease library variability as certain sequences are less likely to be generated. In addition, multiple incorporation would be less efficient when larger modification is applied. More importantly, even though heavily modified dN*TP could be recognized by polymerase, bulky adducts present in template usually blocks DNA polymerase from reading through.[111] In that way, even behaved as potential aptamer, highly modified aptamer lacks the ability to get amplified from PCR which is critical in SELEX process.

For the efficiency and multiple incorporation issue, post-polymerization coupling was introduced. This idea is to pre-introduce a relatively small bio-orthogonal reactive group through modified **dN*TP** and then couple to the modification of interest after the strand is generated (**Figure 1.11**). In that way, the modification would have a relatively small impact for the enzymatic incorporation.

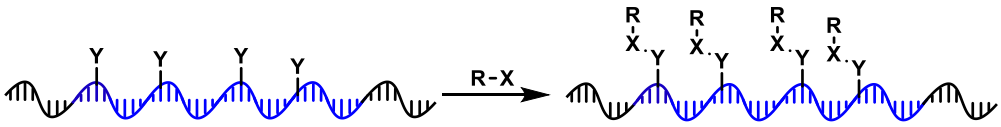


Figure 1.11: Post-polymerization modification.

By replacing dTTP with EdU*TP (C5-ethynyl-2'-dUTP), Mayer et al. was able to introduce bio-orthogonal reactive alkyne onto the template.[112] That was then functionalized through Cu (I) catalyzed cycloaddition with functionalized azide. After selection against target of interest, the desired the strand was then eluted and amplified using EdU*TP through PCR. When one primer used in PCR is 5'-phosphorylated, the undesired strand generated from that primer will get removed using λ-exonuclease digestion. The alkyne strand was then coupled to the modification and continue for evolution. In the characterization study, indole azide was applied in the SELEX against GFP labeled Mouse complement component 3 (C3-GFP). The most potent library member binds to C3-GFP with a K_d value of 18.4nM. Removal or switching out the modification with other functional groups abolish the interaction. Recently, Mayer et al showed a strategy to perform intermolecular cyclization on alkyne modified ssDNA through di-/tri- functionally linker.[113] That allows further stabilization of DNA secondary structure thus increasing the resistance towards nuclease.

With a similar strategy, larger modification could be introduced without worrying about the substrate recognition. It was not about introducing modification but amplifying the modified strand that became the actual problem.

Taking one step back, with the modification presented on the oligonucleotides, it is no longer recognized as nucleic acids by the polymerase. It becomes certain phenotype generated from its genotype (DNA template) using polymerase rather than translation

factors. A strategy that could dissect phenotype and genotype into separate parts would be of great value. As selection is related to phenotype which needs the presence of modification, while amplification is related to genotype which doesn't. An alternative version of an old strategy-DNA display could be helpful in this case. Traditionally, DNA display allows the linkage between DNA and its corresponding potential enzyme after translation (biotin and streptavidin linkage are most commonly applied).[114] With the presence of a DNA tag, the genotype of the selection winner will be enriched through PCR and subjected to the next round.

Compared to proteins or peptides, modified DNA could be linked to its genotype much easier. As modified DNA will be covalently linked to its genotype specifically when using a hairpin structure. Polymerases with strong strand displacement ability[115, 116] allow the insertion of a new strand through the hairpin loop to displace the modified strand (**Figure 1.12**).

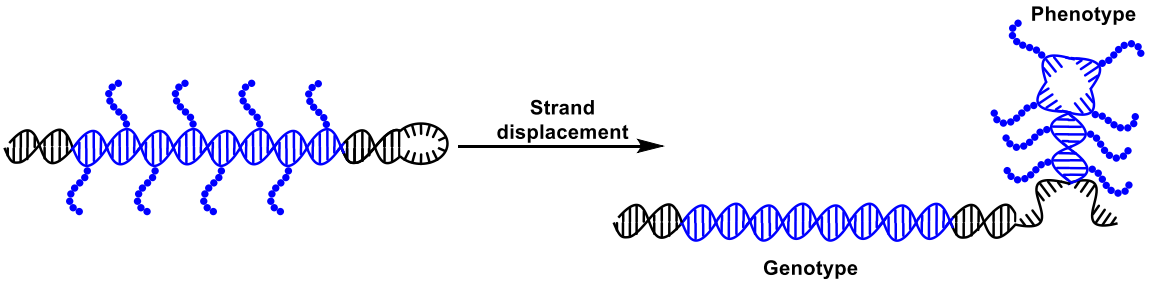


Figure 1.12: DNA-display for heavily modified DNAs.

Krauss et al. developed an approach in evolving sugar molecules ($Mannose_4$) modified ssDNA in mimicking HIV gp120.[117] The sugar molecules were introduced after inserting an alkyne modification onto DNA through polymerase catalyzed primer extension. Sugar molecules are labeled through Cu (I) catalyzed cycloaddition with the azides. Then, a key strand-displacement step comes into play, a short primer is annealed to the loop region

and gets extended through a polymerase with strong displacement ability. In that way, this sugar modified DNA is divided into two parts, dsDNA segment contains the genotype which is ready for PCR amplification and ssDNA segment with sugar modification phenotype for binding towards target of interest (2G12 in this case). Then amplifying the selection winner, perform the whole process again. Following this protocol, sugar modified ssDNA with binding affinity of 260nM towards 2G12 was identified.[117] Removing the modification resulted in complete loss in binding target of interest. After applying the full sugar cluster of $mannose_9$ instead of $mannose_4$ as to mimic the local environment of interface better, aptamer with the K_d of 150nM was found.[118] Further optimization in the SELEX process as elevating selection temperature to remove weak binder resulting in an aptamer with the K_d of 1.7nM[119] which showed the potential of heavily modified aptamers.

With four different nucleosides, the maximum number of modifications is limited to four. Famulok et al. proved that four differently modified dN*TPs were able to get incorporated into a native DNA template.[120] With the potential application in preparing unnatural nucleic acid, the sequence dependence of this enzymatic incorporation still prohibit the applicability of this approach. As it is shown that, GC rich templates were less likely to get incorporated compared to AT rich templates. In addition, polymerase evolution might be required as to accept unnatural substrates, which also builds up the barrier for expanding modification numbers in SELEX.

Functionalized nucleic acid aptamer through modified phosphorylated oligomers

DNA ligase enables the formation of phosphodiester bond between 5'-phosphate end and 3'-hydroxyl group with the presence of ATP. It has been widely applied in biological

system for repairing single stranded breaks, such as ligating okazaki fragments which are generated on the lagging strand for DNA replication.[121]

The process of ligation could be divided into these steps: (*i*) Recognizing the active site which is usually the DNA nick position after proper annealing to the template strand; (*ii*) Adenylating lysine residues at the active site of DNA ligase and releasing pyrophosphate; (*iii*) Transferring AMP on the active site of ligase towards 5'-phosphate strand to replace hydroxyl group with a better leaving group; (*iv*) Ligating 3'-hydroxyl group with the 5'-phoshpate end with a phosphodiester bond.[122] For non-biological applications, DNA ligase was mostly applied in recombinant protein expression. After inserting the DNA sequence for protein of interest through ligation to the multiple cloning region (MCR) on the plasmid, protein of interest is generated through in vitro transcription and translation.[123]

It has been proved that native oligomer substrates are recognized by DNA ligase and ligated onto the DNA template efficiently with sequence specificity.[124] DNA ligase also has the potential for recognizing modified substrates.[125] Even lacking a thorough study on the tolerance of modification, ligase-mediated approach already presented several potential advantages over conventional polymerase mediated approach: (*i*) Decreased stringency on size and nature of modification, as modifications were separated by a certain amount of native nucleotides; (*ii*) Increased coding capacity for different modifications, for instance trimer enables coding for 64 different modifications while pentamer enables 1024; (*iii*) Uniform distribution of modifications, as the each modification is fixed to the specific position based on the modified position of the

oligomers; (*iv*) Potential for longer polymers; (v) Unknown sequencing bias, as the fidelity for introducing modified oligomer has not be investigated much.

Hili and Liu et al. for the first time proposed an approach (**Figure 1.13**) for introducing modification through T4 DNA ligase.[126] 5'-phosphorylated trinucleotides were tested as the substrates for T4 DNA ligase. Although standard ligation failed to generate full-length product, adding PEG 6000 which functions as molecular crowding reagent enables the generation of full-length product. Extending reaction time allows higher conversion to the full-length product. With the presence of 0.1 mg/ml BSA, T4 DNA ligase remained its catalytic activity after 12 hours of the ligation reaction. Further optimization on reagent concentration, reaction temperature enabled efficient incorporation of small molecule modified 5'-phosphorylated trinucleotides through ligase-mediation.

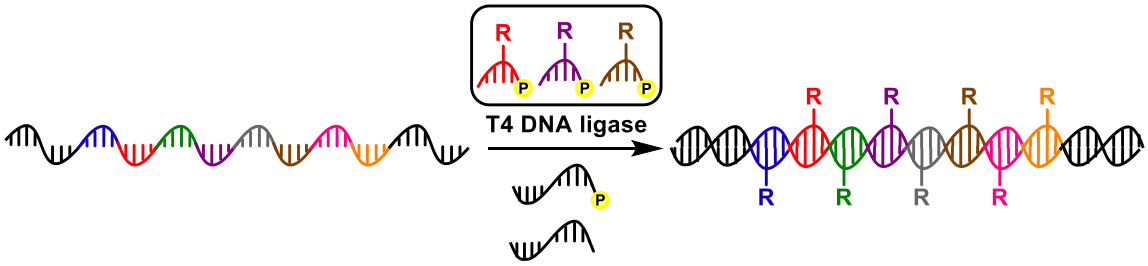


Figure 1.13: Generation of modified DNA through ligase-mediated approach.

According to Hili and Liu et al,[126] T4 DNA ligase allows consecutive incorporation of modified trimers onto DNA template for up to 150 nucleotides (stands for 50 trimer codons). Up to 8 small molecule modifications were introduced through ligase-mediated approach with high efficiency. The chemical diversity was greatly expanded compared to polymerase mediated approach. The sequence fidelity was simply examined by a terminator reaction. In that test, one 5'-phosphorylated trimer was replaced with a trimer of identical sequence without the 5'-phosphate. In that way, once the terminator is

introduced the ligation reaction will be stopped at that point. According to the gel they provide, observed termination product was consistent with the terminator position. That means no readthrough or mis-incorporation occurred, the fidelity of this ligation is high. However, sequencing result will be needed for accurate evaluation for the fidelity. Finally, a mock selection towards carbonic anhydrase II(CAII) beads was performed using trimers modified with various small molecules including a known inhibitor. After four iterative rounds of selection, the positive strand containing the genotype for the inhibitor was enriched by $>2.5 \times 10^7$-fold. That proved the potential of applying ligase-mediated approach for SELEX of modified DNAs with expanded chemical diversity.

1.4 Objectives and dissertation outline

The objectives of this dissertation are as follow: (*i*) Optimize the condition for this ligase-mediated approach as to introduce larger modification. Previously, it has been shown that small molecules modified oligonucleotides were recognized by T4 DNA ligase as substrates. We hypothesized that large peptide modification allows targeting PPIs with high efficiency, while the tolerance to large modifications such as peptides is still unknown right now. (*ii*) Perform mock selection to evaluate the efficiency of this approach in in vitro evolution. The success for iterative rounds of selection rely not only on the choice for starting libraries but also on the robustness of this system. Proper approach to remove undesired product and efficient regeneration of starting material are also critical to the success of the actual SELEX. (*iii*) Evaluate fidelity for this ligase-mediated approach. High fidelity is critical for the correct resembling between genotype and phenotype. Previous evaluation methods (terminator-mediated approach) have

provided an estimation on the sequence specificity while not being a quantitative standard. Sequencing will be required to fulfill this gap. (*iv*) Optimize on the linker-length between modification and oligomer. Longer linker requires higher entropy loss upon binding to the target. Truncation of the linker length with relatively simple approach allows the tuning for identifying aptamer hits with better binding thermodynamics.

The rest of this dissertation is divided into 4 chapters.

Chapter 2 describes the initial work on introducing peptide modifications through T4 DNA ligase. In this case, 5'-phosphorylated pentanucleotides were used as the anti-codon for peptide modification instead of trimers due to higher encoding power (1024 different modifications) and less stringency for the enzyme (less density of modification). An optimization on ATP concentration and template architecture enabled the consecutive incorporation of 8 octa-peptide modified pentamers onto DNA templates. The functional group on peptides were well tolerated by T4 DNA ligase as well including cationic residues, anionic residues, aliphatic residues etc. Finally, a mock selection was performed to evaluate the efficiency of this approach in in vitro evolution, a 190-fold enrichment was observed with one round of selection. This work is published in *J. Am. Chem. Soc.*, **2015**, *137* (34), pp 11191–11196.

Chapter 3 describes the follow up work on introducing peptide modifications through T4 DNA ligase. Lack of the sequence specificity analysis and validated library design limited the application of this ligase-mediated approach. In this chapter, four different ssDNA libraries were first chosen to achieve similar value for ΔG simulation upon folding compared to N_{40} random libraries. That ensured the folding of stable secondary structures. In addition, the melting temperature for octa-peptide modified ssDNA was

measured and compared to native ssDNA as to prove that peptide modifications didn't destabilize the secondary structure. Finally, sequencing result of the ligation product proved the high fidelity of incorporating peptide modified pentanucleotides among the four chosen libraries. One library $(WSWST)_8$ with high yield in generating full-length peptide modified product and high fidelity (98.9%) was chosen as the library for in vitro evolution of potential PPIs inhibitors. This work is published in *Bioconjugate Chem.*, **2017**, *28* (2), pp 314–318.

Chapter 4 describes the further optimization on current ligase-mediated approach for small molecule modification. The lack of tuning ability limited the potential for identifying modified aptamers with improved binding thermodynamics. In this chapter, two strategies will be discussed and compared for truncating the linker length. Direct functionalized phosphoramidite synthesis required multiple synthesis and purifications, while enabled a large variety of modifications. Post-synthesis coupling with the di-amino linker enabled preparation of specific linker with ease and comparable yields, while that coupling reaction has certain limitations. An evaluation on the ligation efficiency and fidelity between linker length was performed. It was proved that truncating the linker slightly resulted in similar yield and fidelity. While removing the linker led to increased sequence bias when incorporating which is not favored for in vitro evolution.

Finally, Chapter 5 provides the summary of the work that is described in this dissertation and a future direction for the possible follow-up work.

1.5 Reference:

1. Tao, Z. F.; Hasvold, L.; Wang, L.; Wang, X.; Petros, A. M.; Park, C. H.; Boghaert, E. R.; Catron, N. D.; Chen, J.; Colman, P. M.; Czabotar, P. E.; Deshayes, K.; Fairbrother, W. J.; Flygare, J. A.; Hymowitz, S. G.; Jin, S.; Judge, R. A.; Koehler, M. F.; Kovar, P. J.; Lessene, G.; Mitten, M. J.; Ndubaku, C. O.; Nimmer, P.; Purkey, H. E.; Oleksijew, A.; Phillips, D. C.; Sleebs, B. E.; Smith, B. J.; Smith, M. L.; Tahir, S. K.; Watson, K. G.; Xiao, Y.; Xue, J.; Zhang, H.; Zobel, K.; Rosenberg, S. H.; Tse, C.; Leverson, J. D.; Elmore, S. W.; Souers, A. J., Discovery of a Potent and Selective BCL-XL Inhibitor with in Vivo Activity. *ACS Med Chem Lett* **2014,** *5* (10), 1088-93.

2. Issaeva, N.; Bozko, P.; Enge, M.; Protopopova, M.; Verhoef, L. G.; Masucci, M.; Pramanik, A.; Selivanova, G., Small molecule RITA binds to p53, blocks p53-HDM-2 interaction and activates p53 function in tumors. *Nat Med* **2004,** *10* (12), 1321-8.

3. Braisted, A. C.; Oslob, J. D.; Delano, W. L.; Hyde, J.; McDowell, R. S.; Waal, N.; Yu, C.; Arkin, M. R.; Raimundo, B. C., Discovery of a potent small molecule IL-2 inhibitor through fragment assembly. *J Am Chem Soc* **2003,** *125* (13), 3714-5.

4. Jones, S.; Thornton, J. M., Principles of protein-protein interactions. *Proc Natl Acad Sci U S A* **1996,** *93* (1), 13-20.

5. Glaser, F.; Morris, R. J.; Najmanovich, R. J.; Laskowski, R. A.; Thornton, J. M., A method for localizing ligand binding pockets in protein structures. *Proteins* **2006,** *62* (2), 479-88.

6. Scott, D. E.; Bayly, A. R.; Abell, C.; Skidmore, J., Small molecules, big targets: drug discovery faces the protein-protein interaction challenge. *Nat Rev Drug Discov* **2016,** *15* (8), 533-50.

7. Robinson, J. A.; Demarco, S.; Gombert, F.; Moehle, K.; Obrecht, D., The design, structures and therapeutic potential of protein epitope mimetics. *Drug Discov Today* **2008,** *13* (21-22), 944-51.

8. Wells, J. A.; McClendon, C. L., Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. *Nature* **2007,** *450* (7172), 1001-9.

9. Congreve, M.; Andrews, S. P.; Dore, A. S.; Hollenstein, K.; Hurrell, E.; Langmead, C. J.; Mason, J. S.; Ng, I. W.; Tehan, B.; Zhukov, A.; Weir, M.; Marshall, F. H., Discovery of 1,2,4-triazine derivatives as adenosine A(2A) antagonists using structure based drug design. *J Med Chem* **2012,** *55* (5), 1898-903.

10. Banaszynski, L. A.; Liu, C. W.; Wandless, T. J., Characterization of the FKBP.rapamycin.FRB ternary complex. *J Am Chem Soc* **2005,** *127* (13), 4715-21.

11. Qing, M.; Yang, F.; Zhang, B.; Zou, G.; Robida, J. M.; Yuan, Z.; Tang, H.; Shi, P. Y., Cyclosporine inhibits flavivirus replication through blocking the interaction between host cyclophilins and viral NS5 protein. *Antimicrob Agents Chemother* **2009,** *53* (8), 3226-35.

12. Clackson, T.; Wells, J. A., A hot spot of binding energy in a hormone-receptor interface. *Science* **1995,** *267* (5196), 383-6.

13. Scott, D. E.; Coyne, A. G.; Hudson, S. A.; Abell, C., Fragment-based approaches in drug discovery and chemical biology. *Biochemistry* **2012,** *51* (25), 4990-5003.

14. Murray, C. W.; Rees, D. C., The rise of fragment-based drug discovery. *Nat Chem* **2009,** *1* (3), 187-92.

15. Petros, A. M.; Dinges, J.; Augeri, D. J.; Baumeister, S. A.; Betebenner, D. A.; Bures, M. G.; Elmore, S. W.; Hajduk, P. J.; Joseph, M. K.; Landis, S. K.; Nettesheim, D.

G.; Rosenberg, S. H.; Shen, W.; Thomas, S.; Wang, X.; Zanze, I.; Zhang, H.; Fesik, S. W., Discovery of a potent inhibitor of the antiapoptotic protein Bcl-xL from NMR and parallel synthesis. *J Med Chem* **2006,** *49* (2), 656-63.

16. Bruncko, M.; Oost, T. K.; Belli, B. A.; Ding, H.; Joseph, M. K.; Kunzer, A.; Martineau, D.; McClellan, W. J.; Mitten, M.; Ng, S. C.; Nimmer, P. M.; Oltersdorf, T.; Park, C. M.; Petros, A. M.; Shoemaker, A. R.; Song, X.; Wang, X.; Wendt, M. D.; Zhang, H.; Fesik, S. W.; Rosenberg, S. H.; Elmore, S. W., Studies leading to potent, dual inhibitors of Bcl-2 and Bcl-xL. *J Med Chem* **2007,** *50* (4), 641-62.

17. Park, C. M.; Bruncko, M.; Adickes, J.; Bauch, J.; Ding, H.; Kunzer, A.; Marsh, K. C.; Nimmer, P.; Shoemaker, A. R.; Song, X.; Tahir, S. K.; Tse, C.; Wang, X.; Wendt, M. D.; Yang, X.; Zhang, H.; Fesik, S. W.; Rosenberg, S. H.; Elmore, S. W., Discovery of an orally bioavailable small molecule inhibitor of prosurvival B-cell lymphoma 2 proteins. *J Med Chem* **2008,** *51* (21), 6902-15.

18. Klages, J.; Coles, M.; Kessler, H., NMR-based screening: a powerful tool in fragment-based drug discovery. *Analyst* **2007,** *132* (7), 693-705.

19. Swayze, E. E.; Jefferson, E. A.; Sannes-Lowery, K. A.; Blyn, L. B.; Risen, L. M.; Arakawa, S.; Osgood, S. A.; Hofstadler, S. A.; Griffey, R. H., SAR by MS: a ligand based technique for drug lead discovery against structured RNA targets. *J Med Chem* **2002,** *45* (18), 3816-9.

20. Hofstadler, S. A.; Sannes-Lowery, K. A., Applications of ESI-MS in drug discovery: interrogation of noncovalent complexes. *Nat Rev Drug Discov* **2006,** *5* (7), 585-95.

21. Vivat Hannah, V.; Atmanene, C.; Zeyer, D.; Van Dorsselaer, A.; Sanglier-Cianferani, S., Native MS: an 'ESI' way to support structure- and fragment-based drug discovery. *Future Med Chem* **2010,** *2* (1), 35-50.

22. Blundell, T. L.; Jhoti, H.; Abell, C., High-throughput crystallography for lead discovery in drug design. *Nat Rev Drug Discov* **2002,** *1* (1), 45-54.

23. Antonysamy, S. S.; Aubol, B.; Blaney, J.; Browner, M. F.; Giannetti, A. M.; Harris, S. F.; Hebert, N.; Hendle, J.; Hopkins, S.; Jefferson, E.; Kissinger, C.; Leveque, V.; Marciano, D.; McGee, E.; Najera, I.; Nolan, B.; Tomimoto, M.; Torres, E.; Wright, T., Fragment-based discovery of hepatitis C virus NS5b RNA polymerase inhibitors. *Bioorg Med Chem Lett* **2008,** *18* (9), 2990-5.

24. Sheng, C.; Dong, G.; Miao, Z.; Zhang, W.; Wang, W., State-of-the-art strategies for targeting protein-protein interactions by small-molecule inhibitors. *Chem Soc Rev* **2015,** *44* (22), 8238-59.

25. Robinson, J. A., Protein epitope mimetics as anti-infectives. *Curr Opin Chem Biol* **2011,** *15* (3), 379-86.

26. (a) Haubner, R.; Schmitt, W.; Hölzemann, G.; Goodman, S. L.; Jonczyk, A.; Kessler, H., Cyclic RGD Peptides Containing β-Turn Mimetics. *J Am Chem Soc* **1996,** *118* (34), 7881-7891; (b) Eguchi, M.; Lee, M. S.; Nakanishi, H.; Stasiak, M.; Lovell, S.; Kahn, M., Solid-Phase Synthesis and Structural Analysis of Bicyclic β-Turn Mimetics Incorporating Functionality at the i to i + 3 Positions. *J Am Chem Soc* **1999,** *121* (51), 12204-12205.

27. Gardner, B.; Nakanishi, H.; Kahn, M., Conformationally constrained nonpeptide β-turn mimetics of enkephalin. *Tetrahedron* **1993,** *49* (17), 3433-3448.

28. Henchey, L. K.; Jochim, A. L.; Arora, P. S., Contemporary strategies for the stabilization of peptides in the alpha-helical conformation. *Curr Opin Chem Biol* **2008,** *12* (6), 692-7.

29. Eichler, J., Peptides as protein binding site mimetics. *Curr Opin Chem Biol* **2008,** *12* (6), 707-13.

30. beta-Hairpin protein epitope mimetic technology in drug discovery. *Drug Discov Today Technol* **2012,** *9* (1), e1-e70.

31. Franke, R.; Doll, C.; Wray, V.; Eichler, J., Loops on loops: generation of complex scaffolded peptides presenting multiple cyclic fragments. *Org Biomol Chem* **2004,** *2* (19), 2847-51.

32. Bullock, B. N.; Jochim, A. L.; Arora, P. S., Assessing helical protein interfaces for inhibitor design. *J Am Chem Soc* **2011,** *133* (36), 14220-3.

33. Kussie, P. H.; Gorina, S.; Marechal, V.; Elenbaas, B.; Moreau, J.; Levine, A. J.; Pavletich, N. P., Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain. *Science* **1996,** *274* (5289), 948-53.

34. Liu, X.; Dai, S.; Zhu, Y.; Marrack, P.; Kappler, J. W., The structure of a Bcl-xL/Bim fragment complex: implications for Bim function. *Immunity* **2003,** *19* (3), 341-52.

35. Czabotar, P. E.; Lee, E. F.; van Delft, M. F.; Day, C. L.; Smith, B. J.; Huang, D. C.; Fairlie, W. D.; Hinds, M. G.; Colman, P. M., Structural insights into the degradation of Mcl-1 induced by BH3 domains. *Proc Natl Acad Sci U S A* **2007,** *104* (15), 6217-22.

36. Jacoby, E., Biphenyls as potential mimetics of protein alpha-helix. *Bioorg Med Chem Lett* **2002,** *12* (6), 891-3.

37. Petros, A. M.; Nettesheim, D. G.; Wang, Y.; Olejniczak, E. T.; Meadows, R. P.; Mack, J.; Swift, K.; Matayoshi, E. D.; Zhang, H.; Thompson, C. B.; Fesik, S. W., Rationale for Bcl-xL/Bad peptide complex formation from structure, mutagenesis, and biophysical studies. *Protein Sci* **2000,** *9* (12), 2528-34.

38. Kutzki, O.; Park, H. S.; Ernst, J. T.; Orner, B. P.; Yin, H.; Hamilton, A. D., Development of a Potent Bcl-xL Antagonist Based on α-Helix Mimicry. *J Am Chem Soc* **2002,** *124* (40), 11838-11839.

39. Ernst, J. T.; Becerril, J.; Park, H. S.; Yin, H.; Hamilton, A. D., Design and application of an alpha-helix-mimetic scaffold based on an oligoamide-foldamer strategy: antagonism of the Bak BH3/Bcl-xL complex. *Angew Chem Int Ed Engl* **2003,** *42* (5), 535-9.

40. Ahn, J.-M.; Han, S.-Y., Facile synthesis of benzamides to mimic an α-helix. *Tetrahedron Lett* **2007,** *48* (20), 3543-3547.

41. Werkhoven, P. R.; Elwakiel, M.; Meuleman, T. J.; Quarles van Ufford, H. C.; Kruijtzer, J. A. W.; Liskamp, R. M. J., Molecular construction of HIV-gp120 discontinuous epitope mimics by assembly of cyclic peptides on an orthogonal alkyne functionalized TAC-scaffold. *Org Biomol Chem* **2016,** *14* (2), 701-710.

42. Brouwer, A. J.; van de Langemheen, H.; Liskamp, R. M. J., Expedient synthesis of a novel asymmetric selectively deprotectable derivative of the ATAC scaffold. *Tetrahedron* **2014,** *70* (26), 4002-4007.

43. Pels, K.; Dickson, P.; An, H.; Kodadek, T., DNA-Compatible Solid-Phase Combinatorial Synthesis of β-Cyanoacrylamides and Related Electrophiles. *ACS Comb Sci* **2018,** *20* (2), 61-69.

44. Gupta, P. B.; Onder, T. T.; Jiang, G.; Tao, K.; Kuperwasser, C.; Weinberg, R. A.; Lander, E. S., Identification of selective inhibitors of cancer stem cells by high-throughput screening. *Cell* **2009,** *138* (4), 645-659.

45. Bandy, T. J.; Brewer, A.; Burns, J. R.; Marth, G.; Nguyen, T.; Stulz, E., DNA as supramolecular scaffold for functional molecules: progress in DNA nanotechnology. *Chem Soc Rev* **2011,** *40* (1), 138-48.

46. Hu, Q.; Li, H.; Wang, L.; Gu, H.; Fan, C., DNA Nanotechnology-Enabled Drug Delivery Systems. *Chem Rev* **2018**.

47. Shendure, J.; Ji, H., Next-generation DNA sequencing. *Nat Biotechnol* **2008,** *26* (10), 1135-45.

48. Richmond, T. J.; Davey, C. A., The structure of DNA in the nucleosome core. *Nature* **2003,** *423* (6936), 145-50.

49. Dickerson, R. E.; Drew, H. R.; Conner, B. N.; Kopka, M. L.; Pjura, P. E., Helix geometry and hydration in A-DNA, B-DNA, and Z-DNA. *Cold Spring Harb Symp Quant Biol* **1983,** *47 Pt 1*, 13-24.

50. Dickerson, R. E.; Klug, A., Base sequence and helix structure variation in B and A DNA. *J Mol Biol* **1983,** *166* (3), 419-441.

51. Whelan, D. R.; Hiscox, T. J.; Rood, J. I.; Bambery, K. R.; McNaughton, D.; Wood, B. R., Detection of an en masse and reversible B- to A-DNA conformational transition in prokaryotes in response to desiccation. *J R Soc Interface* **2014,** *11* (97), 20140454.

52. Fuertes, M. A.; Cepeda, V.; Alonso, C.; Perez, J. M., Molecular mechanisms for the B-Z transition in the example of poly[d(G-C) x d(G-C)] polymers. A critical review. *Chem Rev* **2006,** *106* (6), 2045-64.

53. Roberts, R. W.; Crothers, D. M., Specificity and stringency in DNA triplex formation. *Proc Natl Acad Sci U S A* **1991,** *88* (21), 9397.

54. Risitano, A.; Fox, K. R., Influence of loop size on the stability of intramolecular DNA quadruplexes. *Nucleic Acids Res* **2004,** *32* (8), 2598-2606.

55. Sprinz, K. I.; Tagore, D. M.; Hamilton, A. D., Self-assembly of bivalent protein-binding agents based on oligonucleotide-linked organic fragments. *Bioorg Med Chem Lett* **2005,** *15* (17), 3908-3911.

56. Dumelin, C. E.; Scheuermann, J.; Melkko, S.; Neri, D., Selection of streptavidin binders from a DNA-encoded chemical library. *Bioconjug Chem* **2006,** *17* (2), 366-70.

57. Mannocci, L.; Melkko, S.; Buller, F.; Molnar, I.; Bianke, J. P.; Dumelin, C. E.; Scheuermann, J.; Neri, D., Isolation of potent and specific trypsin inhibitors from a DNA-encoded chemical library. *Bioconjug Chem* **2010,** *21* (10), 1836-41.

58. Williams, B. A.; Diehnelt, C. W.; Belcher, P.; Greving, M.; Woodbury, N. W.; Johnston, S. A.; Chaput, J. C., Creating protein affinity reagents by combining peptide ligands on synthetic DNA scaffolds. *J Am Chem Soc* **2009,** *131* (47), 17233-41.

59. Zhang, J.; Williams, B. A.; Nilsson, M. T.; Chaput, J. C., The evolvability of lead peptides from small library screens. *Chem Commun (Camb)* **2010,** *46* (41), 7778-80.

60. Protozanova, E.; Yakovchuk, P.; Frank-Kamenetskii, M. D., Stacked–Unstacked Equilibrium at the Nick Site of DNA. *J Mol Biol* **2004,** *342* (3), 775-785.

61. Eberhard, H.; Diezmann, F.; Seitz, O., DNA as a molecular ruler: interrogation of a tandem SH2 domain with self-assembled, bivalent DNA-peptide complexes. *Angew Chem Int Ed Engl* **2011,** *50* (18), 4146-50.

62. Marczynke, M.; Groger, K.; Seitz, O., Selective Binders of the Tandem Src Homology 2 Domains in Syk and Zap70 Protein Kinases by DNA-Programmed Spatial Screening. *Bioconjug Chem* **2017,** *28* (9), 2384-2392.

63. Bu, J. Y.; Shaw, A. S.; Chan, A. C., Analysis of the interaction of ZAP-70 and syk protein-tyrosine kinases with the T-cell antigen receptor by plasmon resonance. *Proc Natl Acad Sci U S A* **1995,** *92* (11), 5106-10.

64. Biet, E.; Sun, J. S.; Dutreix, M., Conserved sequence preference in DNA binding among recombination proteins: an effect of ssDNA secondary structure. *Nucleic Acids Res* **1999,** *27* (2), 596-600.

65. Scheibe, C.; Bujotzek, A.; Dernedde, J.; Weber, M.; Seitz, O., DNA-programmed spatial screening of carbohydrate–lectin interactions. *Chem Sci* **2011,** *2* (4), 770-775.

66. Venter, J. C.; Adams, M. D.; Myers, E. W.; Li, P. W.; Mural, R. J.; Sutton, G. G.; Smith, H. O.; Yandell, M.; Evans, C. A.; Holt, R. A.; Gocayne, J. D.; Amanatides, P.; Ballew, R. M.; Huson, D. H.; Wortman, J. R.; Zhang, Q.; Kodira, C. D.; Zheng, X. H.; Chen, L.; Skupski, M.; Subramanian, G.; Thomas, P. D.; Zhang, J.; Gabor Miklos, G. L.; Nelson, C.; Broder, S.; Clark, A. G.; Nadeau, J.; McKusick, V. A.; Zinder, N.; Levine, A. J.; Roberts, R. J.; Simon, M.; Slayman, C.; Hunkapiller, M.; Bolanos, R.; Delcher, A.; Dew, I.; Fasulo, D.; Flanigan, M.; Florea, L.; Halpern, A.; Hannenhalli, S.; Kravitz, S.; Levy, S.; Mobarry, C.; Reinert, K.; Remington, K.; Abu-Threideh, J.; Beasley, E.; Biddick, K.; Bonazzi, V.; Brandon, R.; Cargill, M.; Chandramouliswaran, I.; Charlab, R.;

Chaturvedi, K.; Deng, Z.; Di Francesco, V.; Dunn, P.; Eilbeck, K.; Evangelista, C.; Gabrielian, A. E.; Gan, W.; Ge, W.; Gong, F.; Gu, Z.; Guan, P.; Heiman, T. J.; Higgins, M. E.; Ji, R. R.; Ke, Z.; Ketchum, K. A.; Lai, Z.; Lei, Y.; Li, Z.; Li, J.; Liang, Y.; Lin, X.; Lu, F.; Merkulov, G. V.; Milshina, N.; Moore, H. M.; Naik, A. K.; Narayan, V. A.; Neelam, B.; Nusskern, D.; Rusch, D. B.; Salzberg, S.; Shao, W.; Shue, B.; Sun, J.; Wang, Z.; Wang, A.; Wang, X.; Wang, J.; Wei, M.; Wides, R.; Xiao, C.; Yan, C.; Yao, A.; Ye, J.; Zhan, M.; Zhang, W.; Zhang, H.; Zhao, Q.; Zheng, L.; Zhong, F.; Zhong, W.; Zhu, S.; Zhao, S.; Gilbert, D.; Baumhueter, S.; Spier, G.; Carter, C.; Cravchik, A.; Woodage, T.; Ali, F.; An, H.; Awe, A.; Baldwin, D.; Baden, H.; Barnstead, M.; Barrow, I.; Beeson, K.; Busam, D.; Carver, A.; Center, A.; Cheng, M. L.; Curry, L.; Danaher, S.; Davenport, L.; Desilets, R.; Dietz, S.; Dodson, K.; Doup, L.; Ferriera, S.; Garg, N.; Gluecksmann, A.; Hart, B.; Haynes, J.; Haynes, C.; Heiner, C.; Hladun, S.; Hostin, D.; Houck, J.; Howland, T.; Ibegwam, C.; Johnson, J.; Kalush, F.; Kline, L.; Koduru, S.; Love, A.; Mann, F.; May, D.; McCawley, S.; McIntosh, T.; McMullen, I.; Moy, M.; Moy, L.; Murphy, B.; Nelson, K.; Pfannkoch, C.; Pratts, E.; Puri, V.; Qureshi, H.; Reardon, M.; Rodriguez, R.; Rogers, Y. H.; Romblad, D.; Ruhfel, B.; Scott, R.; Sitter, C.; Smallwood, M.; Stewart, E.; Strong, R.; Suh, E.; Thomas, R.; Tint, N. N.; Tse, S.; Vech, C.; Wang, G.; Wetter, J.; Williams, S.; Williams, M.; Windsor, S.; Winn-Deen, E.; Wolfe, K.; Zaveri, J.; Zaveri, K.; Abril, J. F.; Guigo, R.; Campbell, M. J.; Sjolander, K. V.; Karlak, B.; Kejariwal, A.; Mi, H.; Lazareva, B.; Hatton, T.; Narechania, A.; Diemer, K.; Muruganujan, A.; Guo, N.; Sato, S.; Bafna, V.; Istrail, S.; Lippert, R.; Schwartz, R.; Walenz, B.; Yooseph, S.; Allen, D.; Basu, A.; Baxendale, J.; Blick, L.; Caminha, M.; Carnes-Stine, J.; Caulk, P.; Chiang, Y. H.; Coyne, M.; Dahlke, C.; Mays, A.; Dombroski, M.; Donnelly, M.; Ely, D.; Esparham,

S.; Fosler, C.; Gire, H.; Glanowski, S.; Glasser, K.; Glodek, A.; Gorokhov, M.; Graham, K.; Gropman, B.; Harris, M.; Heil, J.; Henderson, S.; Hoover, J.; Jennings, D.; Jordan, C.; Jordan, J.; Kasha, J.; Kagan, L.; Kraft, C.; Levitsky, A.; Lewis, M.; Liu, X.; Lopez, J.; Ma, D.; Majoros, W.; McDaniel, J.; Murphy, S.; Newman, M.; Nguyen, T.; Nguyen, N.; Nodell, M.; Pan, S.; Peck, J.; Peterson, M.; Rowe, W.; Sanders, R.; Scott, J.; Simpson, M.; Smith, T.; Sprague, A.; Stockwell, T.; Turner, R.; Venter, E.; Wang, M.; Wen, M.; Wu, D.; Wu, M.; Xia, A.; Zandieh, A.; Zhu, X., The sequence of the human genome. *Science* **2001,** *291* (5507), 1304-51.

67. Pley, H. W.; Flaherty, K. M.; McKay, D. B., Three-dimensional structure of a hammerhead ribozyme. *Nature* **1994,** *372*, 68.

68. Tuerk, C.; Gold, L., Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **1990,** *249* (4968), 505-10.

69. Robertson, D. L.; Joyce, G. F., Selection in vitro of an RNA enzyme that specifically cleaves single-stranded DNA. *Nature* **1990,** *344* (6265), 467-8.

70. Ellington, A. D.; Szostak, J. W., In vitro selection of RNA molecules that bind specific ligands. *Nature* **1990,** *346* (6287), 818-22.

71. Vreven, T.; Moal, I. H.; Vangone, A.; Pierce, B. G.; Kastritis, P. L.; Torchala, M.; Chaleil, R.; Jimenez-Garcia, B.; Bates, P. A.; Fernandez-Recio, J.; Bonvin, A. M.; Weng, Z., Updates to the Integrated Protein-Protein Interaction Benchmarks: Docking Benchmark Version 5 and Affinity Benchmark Version 2. *J Mol Biol* **2015,** *427* (19), 3031-41.

72. Fibriansah, G.; Ibarra, K. D.; Ng, T. S.; Smith, S. A.; Tan, J. L.; Lim, X. N.; Ooi, J. S.; Kostyuchenko, V. A.; Wang, J.; de Silva, A. M.; Harris, E.; Crowe, J. E., Jr.; Lok,

S. M., DENGUE VIRUS. Cryo-EM structure of an antibody that neutralizes dengue virus type 2 by locking E protein dimers. *Science* **2015,** *349* (6243), 88-91.

73. Tiller, K. E.; Tessier, P. M., Advances in Antibody Design. *Annu Rev Biomed Eng* **2015,** *17*, 191-216.

74. Zolot, R. S.; Basu, S.; Million, R. P., Antibody-drug conjugates. *Nat Rev Drug Discov* **2013,** *12* (4), 259-60.

75. Ornes, S., Antibody-drug conjugates. *Proc Natl Acad Sci U S A* **2013,** *110* (34), 13695.

76. Jayasena, S. D., Aptamers: an emerging class of molecules that rival antibodies in diagnostics. *Clin Chem* **1999,** *45* (9), 1628-50.

77. Taussig, M. J.; Stoevesandt, O.; Borrebaeck, C. A.; Bradbury, A. R.; Cahill, D.; Cambillau, C.; de Daruvar, A.; Dubel, S.; Eichler, J.; Frank, R.; Gibson, T. J.; Gloriam, D.; Gold, L.; Herberg, F. W.; Hermjakob, H.; Hoheisel, J. D.; Joos, T. O.; Kallioniemi, O.; Koegl, M.; Konthur, Z.; Korn, B.; Kremmer, E.; Krobitsch, S.; Landegren, U.; van der Maarel, S.; McCafferty, J.; Muyldermans, S.; Nygren, P. A.; Palcy, S.; Pluckthun, A.; Polic, B.; Przybylski, M.; Saviranta, P.; Sawyer, A.; Sherman, D. J.; Skerra, A.; Templin, M.; Ueffing, M.; Uhlen, M., ProteomeBinders: planning a European resource of affinity reagents for analysis of the human proteome. *Nat Methods* **2007,** *4* (1), 13-7.

78. Bordeaux, J.; Welsh, A.; Agarwal, S.; Killiam, E.; Baquero, M.; Hanna, J.; Anagnostou, V.; Rimm, D., Antibody validation. *Biotechniques* **2010,** *48* (3), 197-209.

79. Muyldermans, S., Nanobodies: natural single-domain antibodies. *Annu Rev Biochem* **2013,** *82*, 775-97.

80. Klok, H.-A., Peptide/Protein−Synthetic Polymer Conjugates: Quo Vadis. *Macromolecules* **2009,** *42* (21), 7990-8000.

81. Rosanio, G.; Widom, J.; Uhlenbeck, O. C., In vitro selection of DNAs with an increased propensity to form small circles. *Biopolymers* **2015,** *103* (6), 303-20.

82. Avakyan, N.; Greschner, A. A.; Aldaye, F.; Serpell, C. J.; Toader, V.; Petitjean, A.; Sleiman, H. F., Reprogramming the assembly of unmodified DNA with a small molecule. *Nat Chem* **2016,** *8* (4), 368-76.

83. Wang, L.; Ma, W.; Chen, W.; Liu, L.; Ma, W.; Zhu, Y.; Xu, L.; Kuang, H.; Xu, C., An aptamer-based chromatographic strip assay for sensitive toxin semi-quantitative detection. *Biosens Bioelectron* **2011,** *26* (6), 3059-62.

84. Beaucage, S. L.; Iyer, R. P., The synthesis of modified oligonucleotides by the phosphoramidite approach and their applications. *Tetrahedron* **1993,** *49* (28), 6123-6194.

85. Bock, L. C.; Griffin, L. C.; Latham, J. A.; Vermaas, E. H.; Toole, J. J., Selection of single-stranded DNA molecules that bind and inhibit human thrombin. *Nature* **1992,** *355*, 564.

86. Tuerk, C.; MacDougal, S.; Gold, L., RNA pseudoknots that inhibit human immunodeficiency virus type 1 reverse transcriptase. *Proc Natl Acad Sci U S A* **1992,** *89* (15), 6988-6992.

87. Jellinek, D.; Lynott, C. K.; Rifkin, D. B.; Janjic, N., High-Affinity RNA Ligands to Basic Fibroblast Growth Factor Inhibit Receptor Binding. *Proc Natl Acad Sci U S A* **1993,** *90* (23), 11227-11231.

88. Green, L. S.; Jellinek, D.; Bell, C.; Beebe, L. A.; Feistner, B. D.; Gill, S. C.; Jucker, F. M.; Janjić, N., Nuclease-resistant nucleic acid ligands to vascular permeability factor/vascular endothelial growth factor. *Chem Biol* **1995,** *2* (10), 683-695.

89. Klussmann, S.; Nolte, A.; Bald, R.; Erdmann, V. A.; Furste, J. P., Mirror-image RNA that binds D-adenosine. *Nat Biotechnol* **1996,** *14* (9), 1112-5.

90. Jenison, R. D.; Gill, S. C.; Pardi, A.; Polisky, B., High-resolution molecular discrimination by RNA. *Science* **1994,** *263* (5152), 1425-9.

91. Moreira, I. S.; Fernandes, P. A.; Ramos, M. J., Hot spots--a review of the protein-protein interface determinant amino-acid residues. *Proteins* **2007,** *68* (4), 803-12.

92. Parashar, A., Aptamers in Therapeutics. *J Clin Diagn Res* **2016,** *10* (6), BE01-6.

93. Gilar, M.; Bouvier, E. S. P., Purification of crude DNA oligonucleotides by solid-phase extraction and reversed-phase high-performance liquid chromatography. *J Chromatogr A* **2000,** *890* (1), 167-177.

94. Seidu-Larry, S.; Krieg, B.; Hirsch, M.; Helm, M.; Domingo, O., A modified guanosine phosphoramidite for click functionalization of RNA on the sugar edge. *Chem Commun (Camb)* **2012,** *48* (89), 11014-6.

95. Tolle, F.; Rosenthal, M.; Pfeiffer, F.; Mayer, G., Click Reaction on Solid Phase Enables High Fidelity Synthesis of Nucleobase-Modified DNA. *Bioconjug Chem* **2016,** *27* (3), 500-503.

96. Chen, Z.; Lichtor, P. A.; Berliner, A. P.; Chen, J. C.; Liu, D. R., Evolution of sequence-defined highly functionalized nucleic acid polymers. *Nat Chem* **2018,** *10* (4), 420-427.

97. Dignam, J. D.; Lebovitz, R. M.; Roeder, R. G., Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res* **1983,** *11* (5), 1475-1489.

98. Saiki, R. K.; Gelfand, D. H.; Stoffel, S.; Scharf, S. J.; Higuchi, R.; Horn, G. T.; Mullis, K. B.; Erlich, H. A., Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **1988,** *239* (4839), 487.

99. Kunkel, T. A., DNA replication fidelity. *J Biol Chem* **2004,** *279* (17), 16895-8.

100. Watson, J. D.; Crick, F. H. C., Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **1953,** *171*, 737.

101. Giller, G.; Tasara, T.; Angerer, B.; Muhlegger, K.; Amacker, M.; Winter, H., Incorporation of reporter molecule-labeled nucleotides by DNA polymerases. I. Chemical synthesis of various reporter group-labeled 2'-deoxyribonucleoside-5'-triphosphates. *Nucleic Acids Res* **2003,** *31* (10), 2630-5.

102. Bergen, K.; Steck, A. L.; Strutt, S.; Baccaro, A.; Welte, W.; Diederichs, K.; Marx, A., Structures of KlenTaq DNA polymerase caught while incorporating C5-modified pyrimidine and C7-modified 7-deazapurine nucleoside triphosphates. *J Am Chem Soc* **2012,** *134* (29), 11840-3.

103. Latham, J. A.; Johnson, R.; Toole, J. J., The application of a modified nucleotide in aptamer selection: novel thrombin aptamers containing 5-(1-pentynyl)-2'-deoxyuridine. *Nucleic Acids Res* **1994,** *22* (14), 2817-22.

104. Bock, L. C.; Griffin, L. C.; Latham, J. A.; Vermaas, E. H.; Toole, J. J., Selection of single-stranded DNA molecules that bind and inhibit human thrombin. *Nature* **1992,** *355* (6360), 564-6.

105. Gold, L.; Ayers, D.; Bertino, J.; Bock, C.; Bock, A.; Brody, E. N.; Carter, J.; Dalby, A. B.; Eaton, B. E.; Fitzwater, T.; Flather, D.; Forbes, A.; Foreman, T.; Fowler, C.; Gawande, B.; Goss, M.; Gunn, M.; Gupta, S.; Halladay, D.; Heil, J.; Heilig, J.; Hicke, B.; Husar, G.; Janjic, N.; Jarvis, T.; Jennings, S.; Katilius, E.; Keeney, T. R.; Kim, N.; Koch, T. H.; Kraemer, S.; Kroiss, L.; Le, N.; Levine, D.; Lindsey, W.; Lollo, B.; Mayfield, W.; Mehan, M.; Mehler, R.; Nelson, S. K.; Nelson, M.; Nieuwlandt, D.; Nikrad, M.; Ochsner, U.; Ostroff, R. M.; Otis, M.; Parker, T.; Pietrasiewicz, S.; Resnicow, D. I.; Rohloff, J.; Sanders, G.; Sattin, S.; Schneider, D.; Singer, B.; Stanton, M.; Sterkel, A.; Stewart, A.; Stratford, S.; Vaught, J. D.; Vrkljan, M.; Walker, J. J.; Watrobka, M.; Waugh, S.; Weiss, A.; Wilcox, S. K.; Wolfson, A.; Wolk, S. K.; Zhang, C.; Zichi, D., Aptamer-based multiplexed proteomic technology for biomarker discovery. *PLoS One* **2010,** *5* (12), e15004.

106. Vaught, J. D.; Bock, C.; Carter, J.; Fitzwater, T.; Otis, M.; Schneider, D.; Rolando, J.; Waugh, S.; Wilcox, S. K.; Eaton, B. E., Expanding the chemistry of DNA for in vitro selection. *J Am Chem Soc* **2010,** *132* (12), 4141-51.

107. Gawande, B. N.; Rohloff, J. C.; Carter, J. D.; von Carlowitz, I.; Zhang, C.; Schneider, D. J.; Janjic, N., Selection of DNA aptamers with two modified bases. *Proc Natl Acad Sci U S A* **2017,** *114* (11), 2898-2903.

108. Baccaro, A.; Steck, A. L.; Marx, A., Barcoded nucleotides. *Angew Chem Int Ed Engl* **2012,** *51* (1), 254-7.

109. Welter, M.; Verga, D.; Marx, A., Sequence-Specific Incorporation of Enzyme-Nucleotide Chimera by DNA Polymerases. *Angew Chem Int Ed Engl* **2016,** *55* (34), 10131-5.

110. Balintová, J.; Welter, M.; Marx, A., Antibody–nucleotide conjugate as a substrate for DNA polymerases. *Chem Sci* **2018**.

111. Hogg, M.; Wallace, S. S.; Doublie, S., Bumps in the road: how replicative DNA polymerases see DNA damage. *Curr Opin Struct Biol* **2005,** *15* (1), 86-93.

112. Tolle, F.; Brandle, G. M.; Matzner, D.; Mayer, G., A Versatile Approach Towards Nucleobase-Modified Aptamers. *Angew Chem Int Ed Engl* **2015,** *54* (37), 10971-4.

113. Seyfried, P.; Eiden, L.; Grebenovsky, N.; Mayer, G.; Heckel, A., Photo-Tethers for the (Multi-)Cyclic, Conformational Caging of Long Oligonucleotides. *Angew Chem Int Ed Engl* **2017,** *56* (1), 359-363.

114. Yonezawa, M.; Doi, N.; Kawahashi, Y.; Higashinakagawa, T.; Yanagawa, H., DNA display for in vitro selection of diverse peptide libraries. *Nucleic Acids Res* **2003,** *31* (19), e118-e118.

115. Notomi, T.; Okayama, H.; Masubuchi, H.; Yonekawa, T.; Watanabe, K.; Amino, N.; Hase, T., Loop-mediated isothermal amplification of DNA. *Nucleic Acids Res* **2000,** *28* (12), E63.

116. Blanco, L.; Bernad, A.; Lazaro, J. M.; Martin, G.; Garmendia, C.; Salas, M., Highly efficient DNA synthesis by the phage phi 29 DNA polymerase. Symmetrical mode of DNA replication. *J Biol Chem* **1989,** *264* (15), 8935-40.

117. MacPherson, I. S.; Temme, J. S.; Habeshian, S.; Felczak, K.; Pankiewicz, K.; Hedstrom, L.; Krauss, I. J., Multivalent Glycocluster Design through Directed Evolution. *Angew Chem Int Ed Engl* **2011,** *50* (47), 11238-11242.

118. Temme, J. S.; Drzyzga, M. G.; MacPherson, I. S.; Krauss, I. J., Directed Evolution of 2G12-Targeted Nonamannose Glycoclusters by SELMA. *Chem Eur J* **2013,** *19* (51), 17291-17295.

119. Temme, J. S.; MacPherson, I. S.; DeCourcey, J. F.; Krauss, I. J., High temperature SELMA: evolution of DNA-supported oligomannose clusters which are tightly recognized by HIV bnAb 2G12. *J Am Chem Soc* **2014,** *136* (5), 1726-9.

120. Jager, S.; Rasched, G.; Kornreich-Leshem, H.; Engeser, M.; Thum, O.; Famulok, M., A versatile toolbox for variable DNA functionalization at high density. *J Am Chem Soc* **2005,** *127* (43), 15071-82.

121. Okazaki, R.; Okazaki, T.; Sakabe, K.; Sugimoto, K., Mechanism of DNA replication possible discontinuity of DNA chain growth. *Jpn J Med Sci Biol* **1967,** *20* (3), 255-60.

122. Lehman, I. R., DNA ligase: structure, mechanism, and function. *Science* **1974,** *186* (4166), 790-7.

123. Rosano, G. L.; Ceccarelli, E. A., Recombinant protein expression in Escherichia coli: advances and challenges. *Front Microbiol* **2014,** *5*, 172.

124. Kaczorowski, T.; Szybalski, W., Assembly of 18-nucleotide primers by ligation of three hexamers: sequencing of large genomes by primer walking. *Anal Biochem* **1994,** *221* (1), 127-35.

125. Liang, X.; Fujioka, K.; Asanuma, H., Nick sealing by T4 DNA ligase on a modified DNA template: tethering a functional molecule on D-threoninol. *Chemistry* **2011,** *17* (37), 10388-96.

126. Hili, R.; Niu, J.; Liu, D. R., DNA ligase-mediated translation of DNA into densely functionalized nucleic acid polymers. *J Am Chem Soc* **2013,** *135* (1), 98-101.

CHAPTER 2

Sequence-defined scaffolding of peptides on nucleic acid polymers[1]

1. Guo, C.; Watkins, C. P.; Hili, R. *J Am Chem Soc* **2015,** *137* (34), 11191-6.

2.1 Abstract

We have developed a method for the T4 DNA ligase-catalyzed DNA-templated polymerization of 5'-phosphorylated pentanucleotides containing peptide fragments. The polymerization proceeds sequence-specifically to generate DNA-scaffolded peptides in excellent yields. The method has been shown to tolerate peptides ranging from 2-8 amino acids in length with a wide variety of functionality. We validated the capabilities of this system in a mock selection for the enrichment of His-tagged DNA-scaffolded peptide phenotypes from a library, which exhibited a190-fold enrichment after one round of selection. This strategy demonstrates a promising new approach to enable the generation and in vitro selection of high-affinity reagents based upon ssDNA scaffolding of peptide fragments.

2.2 Introduction

Nature uses multi-valency – the sum of low-affinity molecular interactions – to achieve specific and high-affinity molecular recognition.[1] Chemists have strived to apply this concept toward the development of novel high-affinity reagents to address critical needs in biomedical research.[2-4] Not surprisingly, the molecular scaffold is crucial for achieving precise multivalent display of ligands. Thus, the ability of nucleic acids to predictably and reproducibly form tertiary structures makes them an excellent scaffold for displaying ligands in a rigid and predefined spatial configuration.[5-8] Indeed, DNA has been used to scaffold known ligands to generate multivalent high-affinity reagents that rival the properties of traditional antibodies.[9-14]Multivalent display of bioactive oligomers, such as peptides and glycans, has proved particularly effective; however, current approaches

require existing knowledge of ligand binding and typically depend on homo-multi-valency. To fully realize the potential of DNA scaffolding for the development of novel high-affinity reagents, technologies that enable the sequence-defined display of a library of oligomeric ligands along a library of ssDNA at high densities are needed. This would enable the directed evolution of both the ssDNA scaffold architecture and the identity of the displayed oligomers to occur concomitantly in order to optimize recognition of molecular targets.

Inspired by the antigen-binding loops within the complementarity-determining regions of immunoglobulins,[15] we hypothesized that single-stranded nucleic acid polymers decorated in a sequence-defined manner with short peptide fragments could mimic the surface of proteins and serve as a new class of high-affinity reagents (**Figure 2.1a**). The single-stranded nucleic acid component would function as a core scaffold to display multiple unique peptides, generating the hetero-multi-valency akin to hot spots presented at protein-protein interfaces.[16] The T4 DNA ligase-catalyzed DNA-templated polymerization of 5'-phosphorylated trinucleotides provides a platform for the incorporation of multiple small ligands throughout a ssDNA polymer.[17] We reasoned that this approach could be optimized for the polymerization of 5'-phosphorylated oligonucleotides containing peptide fragments to readily assemble the desired DNA-scaffolded peptides (**Figure 2.1b**).

Herein, we report the development of a T4 DNA ligase-catalyzed DNA-templated polymerization of peptide-modified pentanucleotides that enables the sequence-defined scaffolding of peptide fragments on nucleic acid polymers. We examined and optimized this approach to achieve high-fidelity polymerizations with excellent efficiencies and

peptide substrate scope. In addition, we validate the capabilities of the developed polymerization system by integrating it within a mock in vitro selection of functional DNA-scaffolded peptides. The findings from this work advance the field of DNA-templated polymerization and lay the foundation for the evolution of synthetic polymers for molecular recognition and catalysis that are based upon the hetero-multivalent-scaffolding of multiple unique peptide fragments on ssDNA.

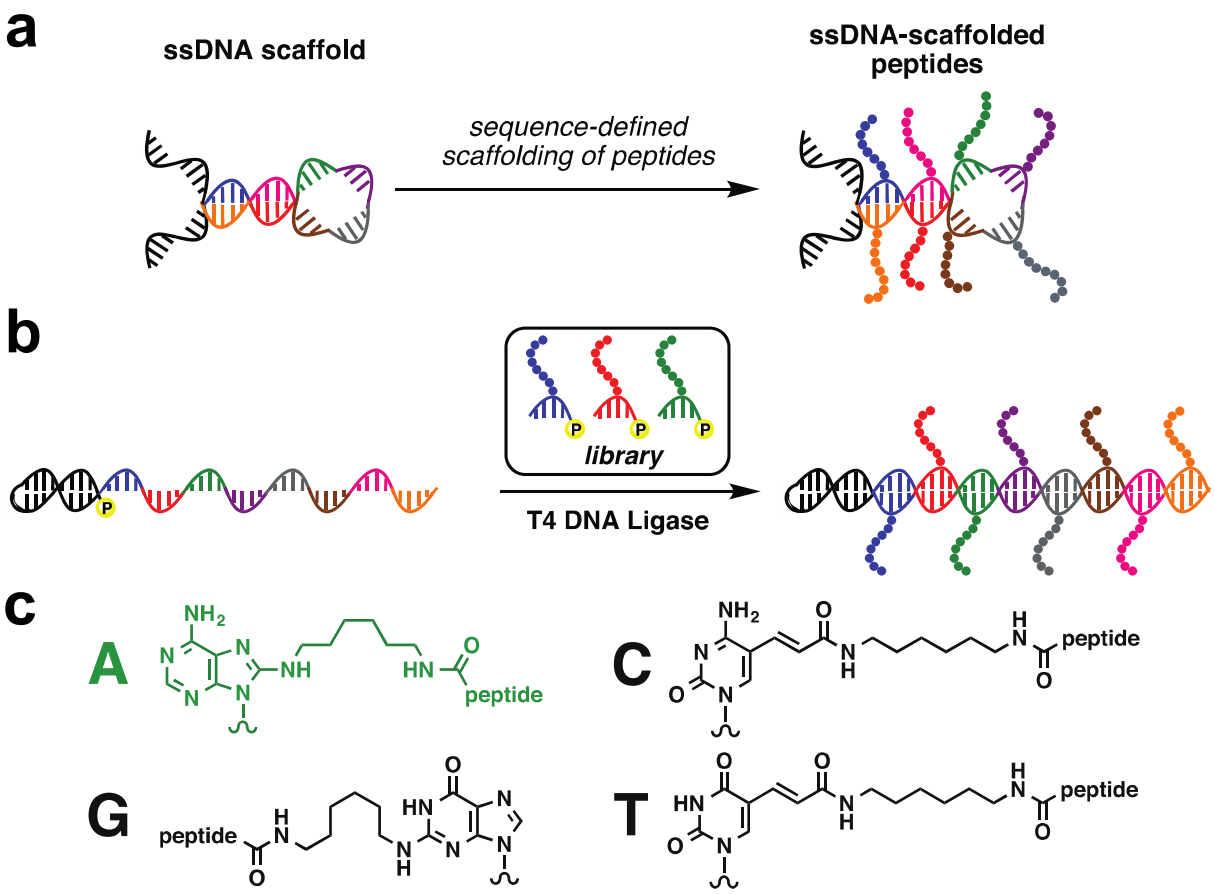


Figure 2.1: Generation of DNA-scaffolded peptides. (**a**) ssDNA-scaffolded peptides. (**b**) Sequence-defined T4 DNA ligase-catalyzed DNA-templated polymerization of peptide-modified oligonucleotides toward the synthesis of ssDNA-scaffolded peptides. (**c**) Peptide modification via commercial amino-modified nucleobases.

2.3 Results and discussions

Polymerization optimization for dipeptide-modified pentanucleotides

We explored the use of peptide-modified pentanucleotides as minimal building blocks for T4 DNA ligase-mediated templated polymerization. We speculated that this oligonucleotide length would provide advantages over shorter codon lengths, including access to longer peptide modifications and a larger codon set, while still providing the desired density of scaffolding and hybridization specificity observed with shorter oligonucleotides.[18] First, we characterized and optimized the ability of T4 DNA ligase to polymerize the 5'-phosphorylated pentanucleotide ACTCT modified with the dipeptide Ac-Phe-Gly at the first nucleobase position via an amino-modified dA (**Figure 2.1c**). The pentanucleotide building blocks were readily synthesized using automated oligonucleotide synthesis with commercially available amine-modified nucleoside phosphoramidites and 5'-phosphorylation reagents. The amine groups were used to install the dipeptide using well-established amide bond-forming chemistry.[19]The desired dipeptide-modified 5'-phosphorylated pentanucleotides were furnished in high yield following purification by reverse-phase HPLC.

An initial optimization screen for polymerization of the dipeptide building block was performed along a DNA template comprising a 5'-phosphorylated hairpin as the extension site followed by eight consecutive repeats of the corresponding pentanucleotide codon (**Figure 2.2**). Templates contained an additional 3'-end nucleotide to preclude undesired blunt-end ligation. The initial polymerization conditions used were adapted from the trinucleotide polymerization system,[17] which contained 10% PEG 6000 as a molecular crowding reagent [20] and 1 mM ATP for 24 h at 25 °C. These conditions

resulted in incomplete polymerization, which was evidenced by a ladder of eight bands up to full-length product seen by denaturing PAGE analysis (**Figure 2.2**, [ATP] = 1000 μM).

We hypothesized that T4 DNA ligase was inefficient at catalyzing the formation of the phosphodiester bond between the 5'-adenylated template and the 3'-hydroxyl group of transiently hybridized peptide-modified pentanucleotides. Under such conditions, the ligase would eventually dissociate from the template and re-adenylate itself in solution. Since adenylated ligase catalyzes the adenylation of the 5'-phosphorylated templates, while the un-adenylated ligase catalyzes the phosphodiester bond formation,[21] we reasoned that the standard high concentration of ATP was inhibiting the polymerization by shifting the equilibrium of the reaction to a mixture of 5'-adenylated template and adenylated ligase, effectively shutting down the polymerization. Indeed, when the ATP concentration was decreased to 25 μM, polymerization of the dipeptide-modified pentanucelotide proceeded smoothly to full-length product (**Figure 2.2**).

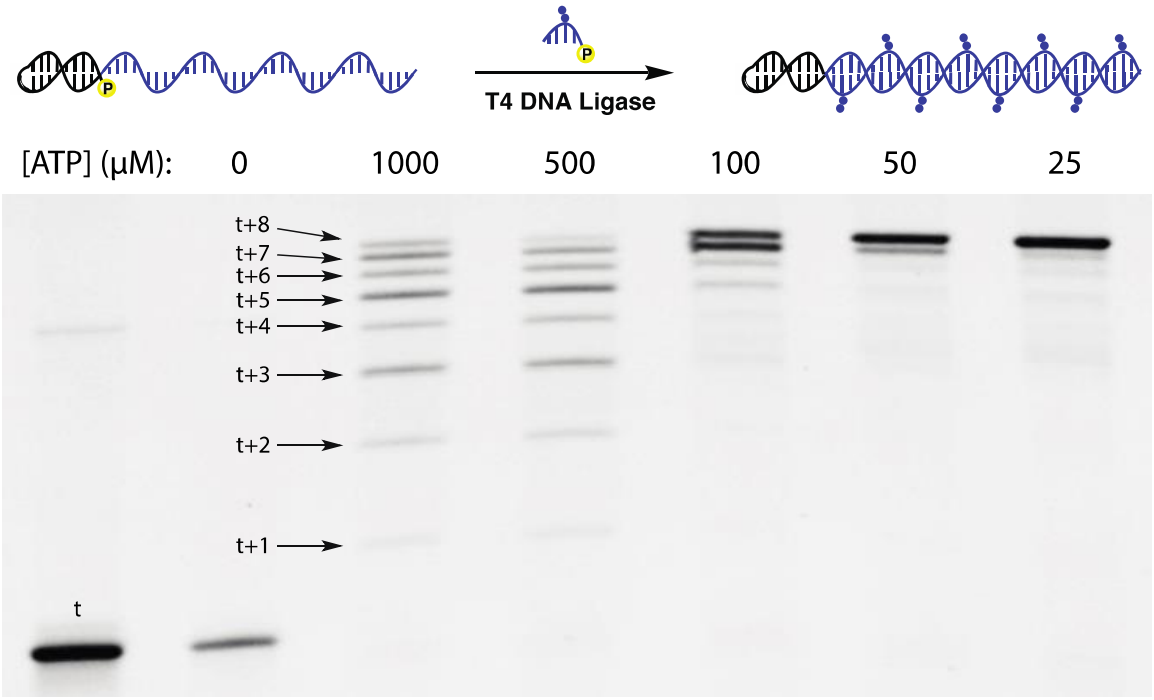


Figure 2.2: Optimization of ATP concentration for the T4 DNA ligase-mediate polymerization of Ac-Phe-Gly-modified pentanucleotide 5'P-ACTCT. Conditions: 1μM template, 4 μM/codon modified 5'-phosphorylated pentanucleotide, 20 U/μL T4 DNA ligase, 10% PEG 6000, and variable ATP for 24 h at 25 °C. t = template**.**

We next performed a positional scan for the modification site on the pentanucleotide building block. A series of pentanucleotides were synthesized with the dipeptide Ac-Phe-Gly at one of the five possible positions via an amino-modified dA nucleotide (**Figure 2.3**). PAGE analysis of the polymerization revealed that all modified positions except position five were tolerated by T4 DNA ligase; modification at position five resulted in extensive laddering up to full-length product. We also studied the nucleobase dependence at position one when modified with Ac-Phe-Gly via amino-modified dA, dT, dC, or dG (See Supporting Information, **Figure S2.2**). Surprisingly, only peptide modification of dA at position one was tolerated; however, a more thorough nucleobase screen across all

positions of the pentanucleotide is required to fully understand the substrate tolerance of T4 DNA ligase.

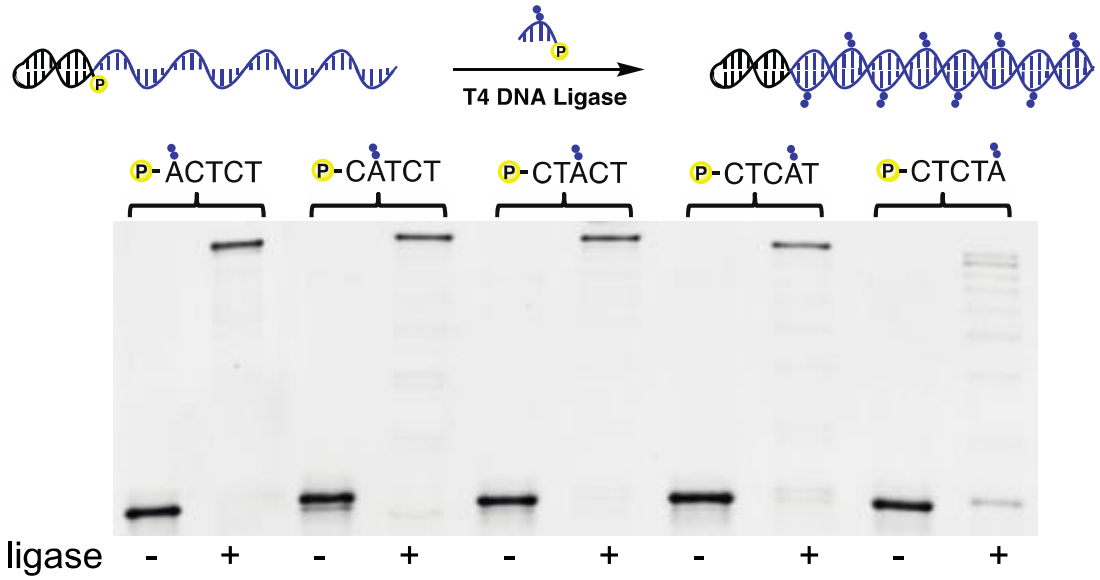


Figure 2.3: Positional scan of the modification site along the pentanucleotide. The dipeptide used was Ac-Phe-Gly, which was conjugated to the pentanucleotide via a C6 amino dA modifier through standard amide coupling

Polymerization of pentanucleotides with longer peptide modifications

Building from the above optimization studies, we sought to expand the scope of the polymerization process to include longer peptides. 5'-phosphorylated pentanucleotide building block ACTCT was modified via an amino-modified dA with peptides ranging from two to eight amino acids in length and polymerized along a homo-octameric template containing a 5'-phosphorylated hairpin (**Figure 2.4**).

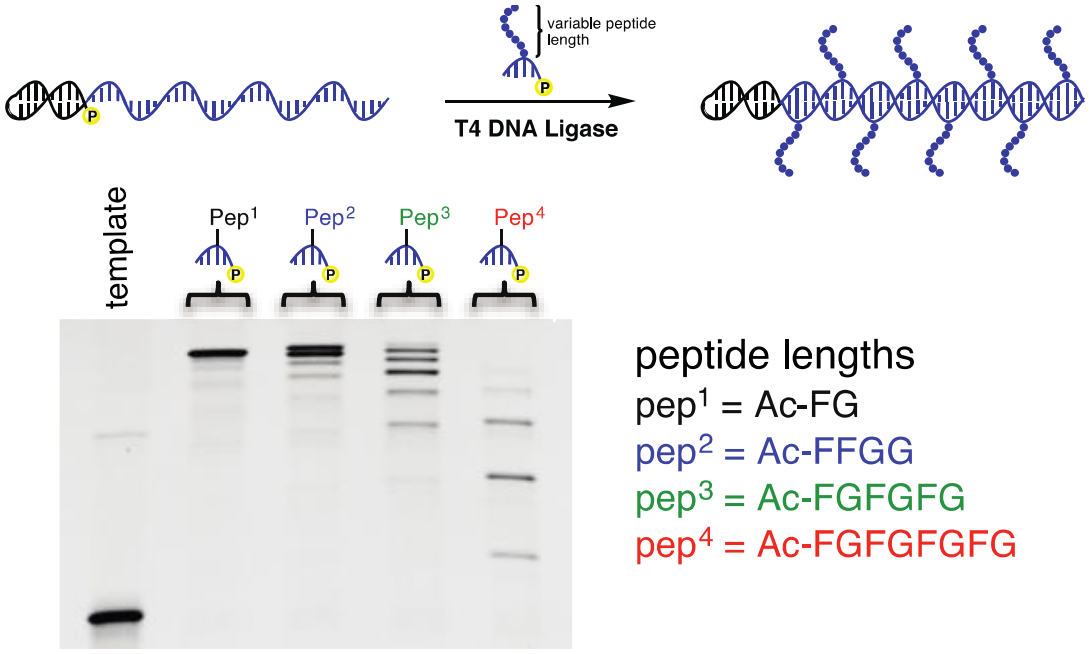


Figure 2.4: Denaturing PAGE analysis of polymerizations of pentanucleotides containing peptide modifications of increasing length along a 5'-hairpin DNA template. The sequence for the various peptide length were presented above, all of them were conjugated to the pentanucleotides via C6 amino dA modified though standard amide coupling

We observed a steady decline in polymerization efficiency as peptide length increased, suggesting that T4 DNA ligase was sensitive to the size of the modification and that further optimization was required. We hypothesized that different template architectures, especially those that enable simultaneous extension from both a 3'-primer and a 5'-primer might enable more efficient ligase-catalyzed polymerization. We chose to explore different template architectures using the most challenging substrate, the octapeptide-modified pentanucleotide (**Figure 2.5**).

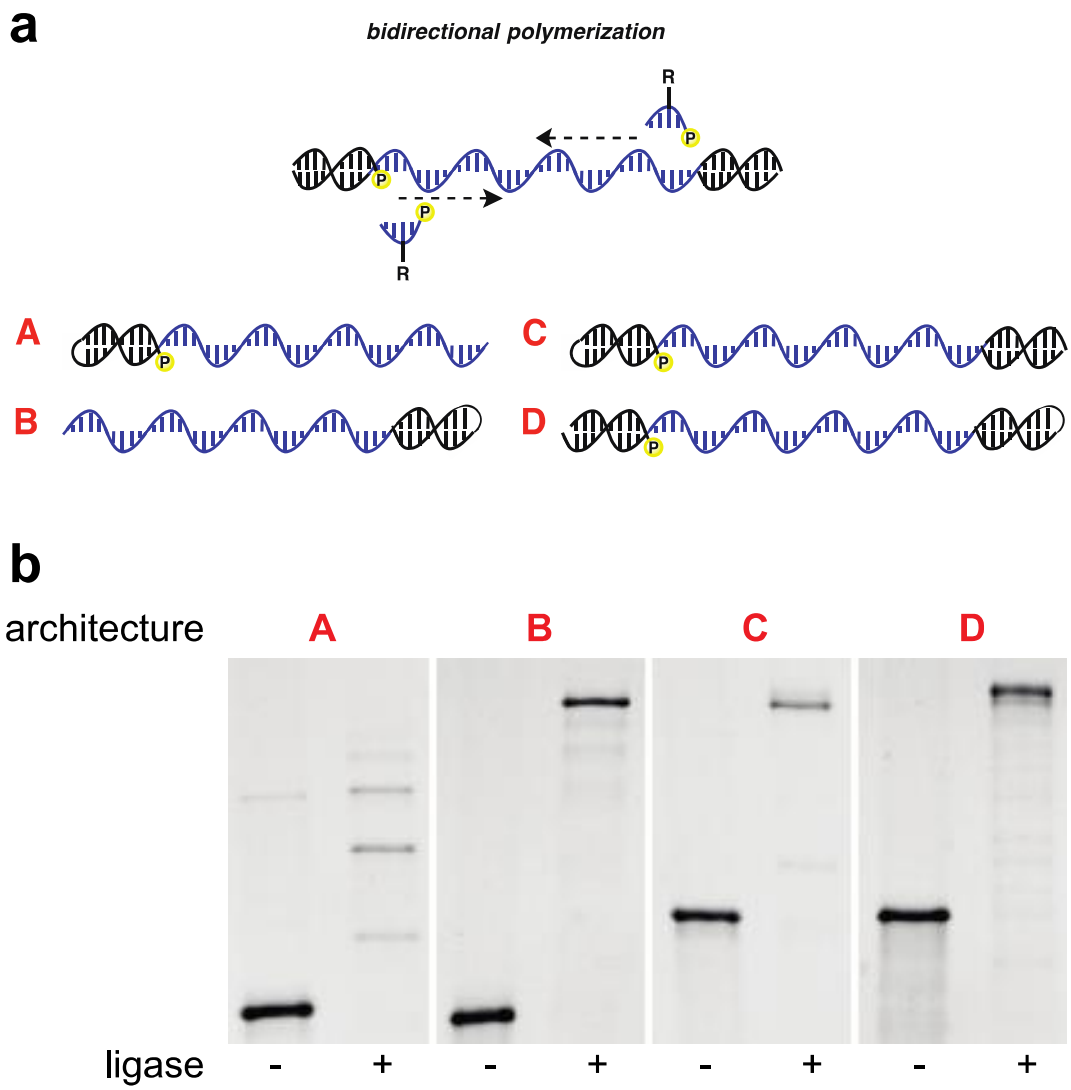


Figure 2.5: Influence of template architecture on the polymerization of pentanucleotides modified with Ac-FGFGFGFG. (**a**) Examined template architectures. (**b**) Denaturing PAGE analysis for 24 h polymerization of octapeptide-modified pentanucleotide on different template architectures.

When comparing between polymerizations of the octapeptide-modified pentanucleotide along templates containing either the 3'-hairpin or 5'-hairpin architectures, we saw a dramatic change in efficiency (**Figure 2.5b**). Polymerizations that extended from the 3'-hairpin proceeded with excellent conversion into full-length product, while extensions

from the 5'-hairpin resulted in a laddering of up to four pentanucleotide incorporations. Polymerizations on template architectures that contained both a hairpin and a primer-binding site for bidirectional extension proceeded efficiently regardless of the location of the hairpin. Importantly, 3'-hairpin templates enable displacement and display of modified single-stranded nucleic acids used during in vitro selections for molecular recognition. [9, 11, 22]

We next challenged the polymerization system with octapeptides containing a diverse set of amino acids including hydrophobic, hydrophilic, and charged residues (**Figure 2.6**). The cationic peptide Ac-WSKGRGSC and the anionic peptide Ac-EGYDTGSC were separately conjugated to the 5'-phosphorylated pentanucleotide ACTCT (**Figure 2.6b**). The conjugation was achieved by coupling the amino group of the C6 amino dA to the C-terminal cysteine residue of the peptide via a succinimidyl 4-(N-maleimidomethyl) cyclohexane-1-carboxylate (SMCC) linker. Polymerizations were performed on the corresponding 3'-hairpin templates using the optimized reaction conditions for 24 h (**Figure 2.6a**). Highly efficient polymerization into full-length product was observed for both peptide-containing substrates (**Figure 2.6c**). These data suggest that the developed polymerization system can accommodate a broad scope of peptide modifications.

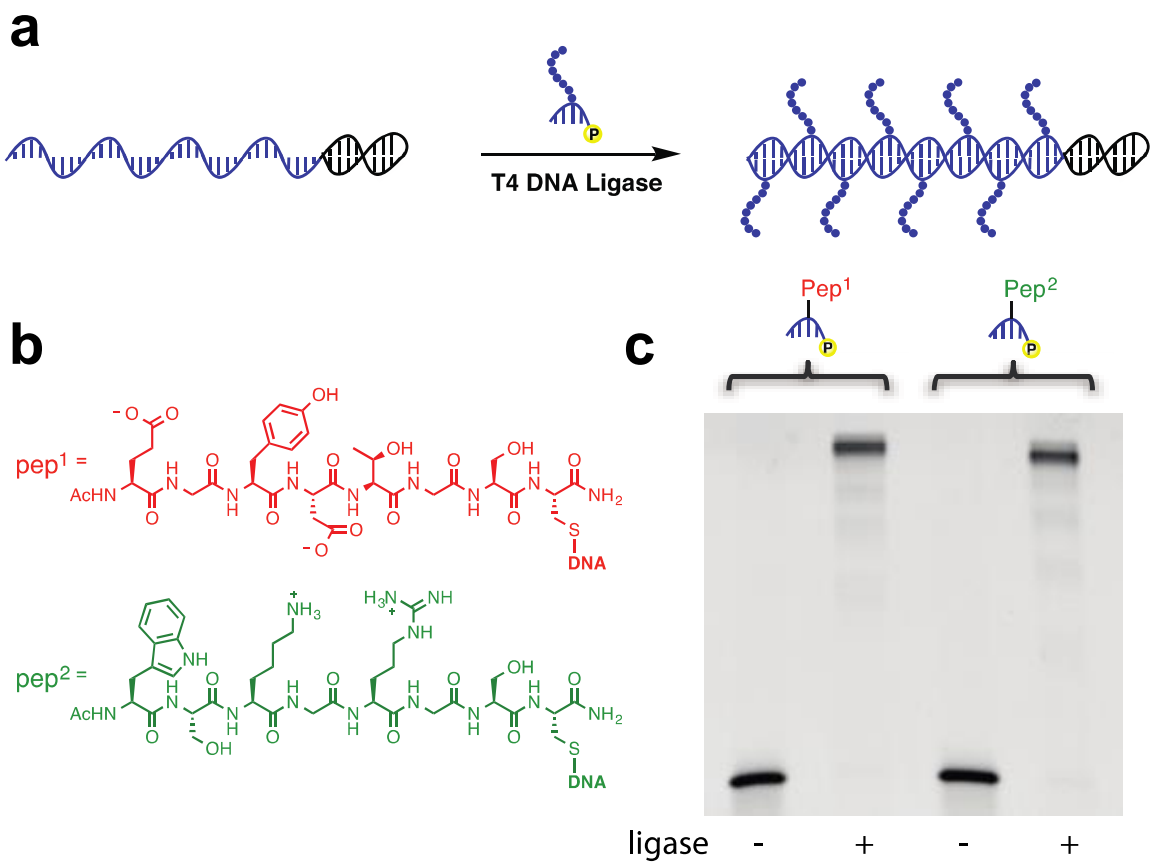


Figure 2.6: Scaffolding of densely functionalized peptides on DNA. (**a**) Optimized 3'-hairpin polymerization strategy. (**b**) Peptides used for polymerization. (**c**) Denaturing PAGE analysis of 24 h polymerization.

Sequence specificity of polymerization

To evaluate the sequence specificity of peptide-modified pentanucleotide incorporation along a 3'-hairpin template, we used a polymerization inhibitor strategy. Since T4 DNA ligase requires 5'-phosphorylated pentanucleotides for the polymerization, the addition of pentanucleotides that lack a 5'-phosphate should inhibit polymerization and result in truncation products. If T4 DNA ligase is highly specific, then polymerization should terminate at the codon that specifies the non-phosphorylated pentanucleotide. If sequence specificity is poor, then polymerization should generate polymers of undesired lengths

either by the mis-incorporation of a non-terminator substrate opposite the terminator codon, or by nonspecific inhibition of polymerization by terminators at non-terminator codons.

We used a hetero-tetrameric template containing four codons with a GC-content ranging from 0-40 % and a 3'-hairpin as the extension site (**Figure 2.7**). Three 5'-phosphorylated pentanucleotides linked to the dipeptide Ac-Phe-Gly and one peptide--modified pentanucleotide lacking a 5'-phosphate were polymerized along the template and the reaction products were analyzed by denaturing PAGE. One product at the anticipated molecular weight was observed for each terminator reaction, indicating that incorporation of all four peptide-modified pentanucleotides proceeds with a high degree of specificity. It is important to highlight that non-phosphorylated pentanucleotides can only inhibit polymerization, as they cannot be incorporated into the polymer; despite a 24 h incubation, and a four-fold excess of phosphorylated pentanucleotides, read-through of the termination codon was not observed.

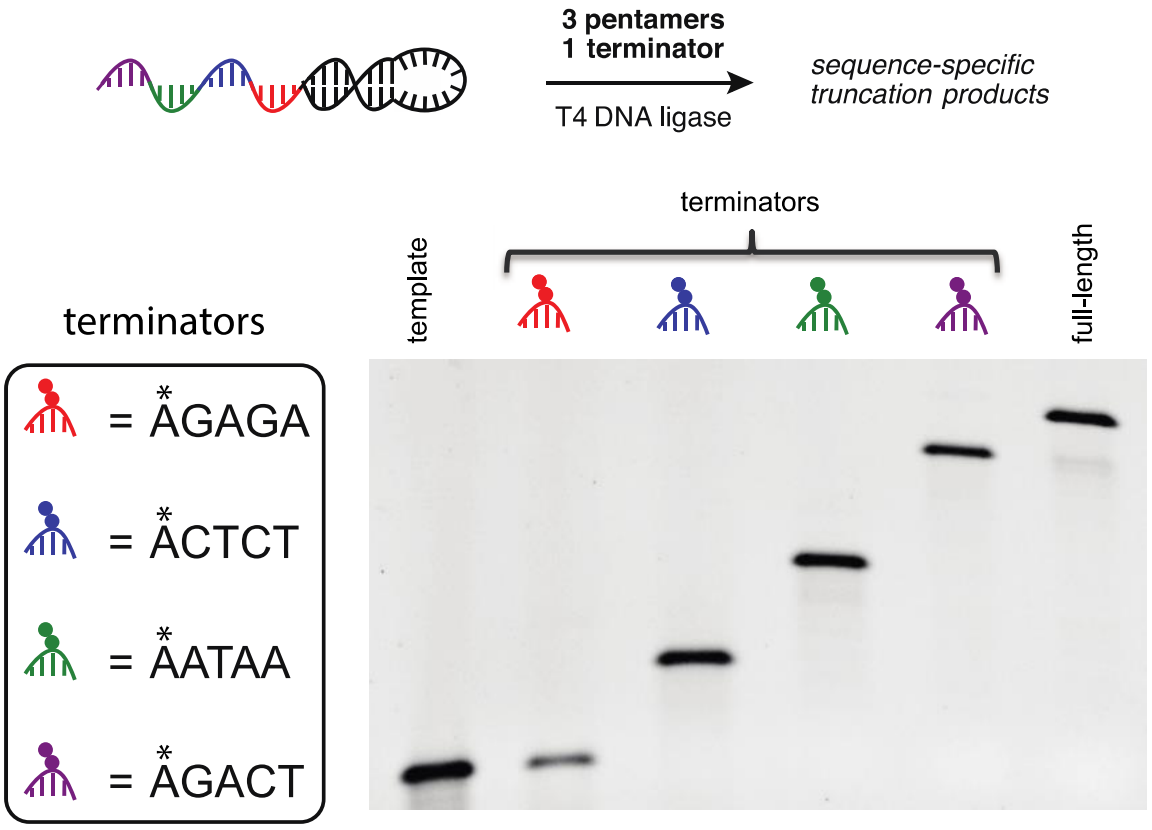


Figure 2.7: Sequence specificity of T4 DNA ligase-mediated templated polymerization of dipeptide-modified pentanucleotides. Asterisks designate sites of peptide modification. The dipeptide used was Ac-Phe-Gly, which was conjugated to both terminator and non-terminator pentanucleotides via a C6 amino dA modifier through standard amide coupling.

In vitro selection *of a hexahistidine phenotype from a library of DNA-scaffolded peptides*

Encouraged by the efficiency and fidelity of the T4 DNA ligase-catalyzed DNA-templated polymerization of peptide-modified pentanucleotides, we sought to test its performance within the context of an in vitro selection cycle to enrich a known binder from a library of DNA-scaffolded peptides. Since the presence of the peptide fragments could potentially interfere with the PCR amplification, we adapted a DNA-display selection approach,[9, 11, 22] which obviates the need for amplifying modified DNA by

covalently linking genotype to phenotype (**Figure 2.8a**). As a model selection, we chose a positive control genotype that encoded for a DNA-scaffolded peptide phenotype containing one instance of a hexahistidine peptide, which would enable it to survive a selection pressure based upon binding to Co^{2+}magnetic particles. For the selection system, we used the four pentanucleotide sequences that we previously demonstrated to be highly specific in the polymerization process, and modified them with four different peptide fragments, including the hexahistidine tag (**Figure 2.8b**). We designed a DNA template (POS) comprising a 3'-hairpin as the extension site followed by seven codons that encoded for pentanucleotides **2**-**4**, and an eighth codon that encoded for the hexahistidine pentanucleotide **1**; the reading frame was followed by a primer-binding site. A library of DNA templated (LIB) was also prepared, whereby the reading frame comprised only codons that encode for pentanucleotides **2**-**4**. The POS template had a unique *XbaI* digest site, so that its enrichment versus LIB could be monitored by restriction digest and PAGE analysis. The POS template was diluted 1000-fold into LIB, and subjected to one round of polymerization, primer extension, selection. The surviving genotypes were amplified by PCR and subjected to digestion by *XbaI* and the digestion results were analyzed by non-denaturing PAGE. Over the one round of selection, the POS genotype was enriched 190-fold (**Figure 2.8c**), demonstrating the capability of the system to support the in vitro selection of DNA-scaffolded peptides against molecular targets.

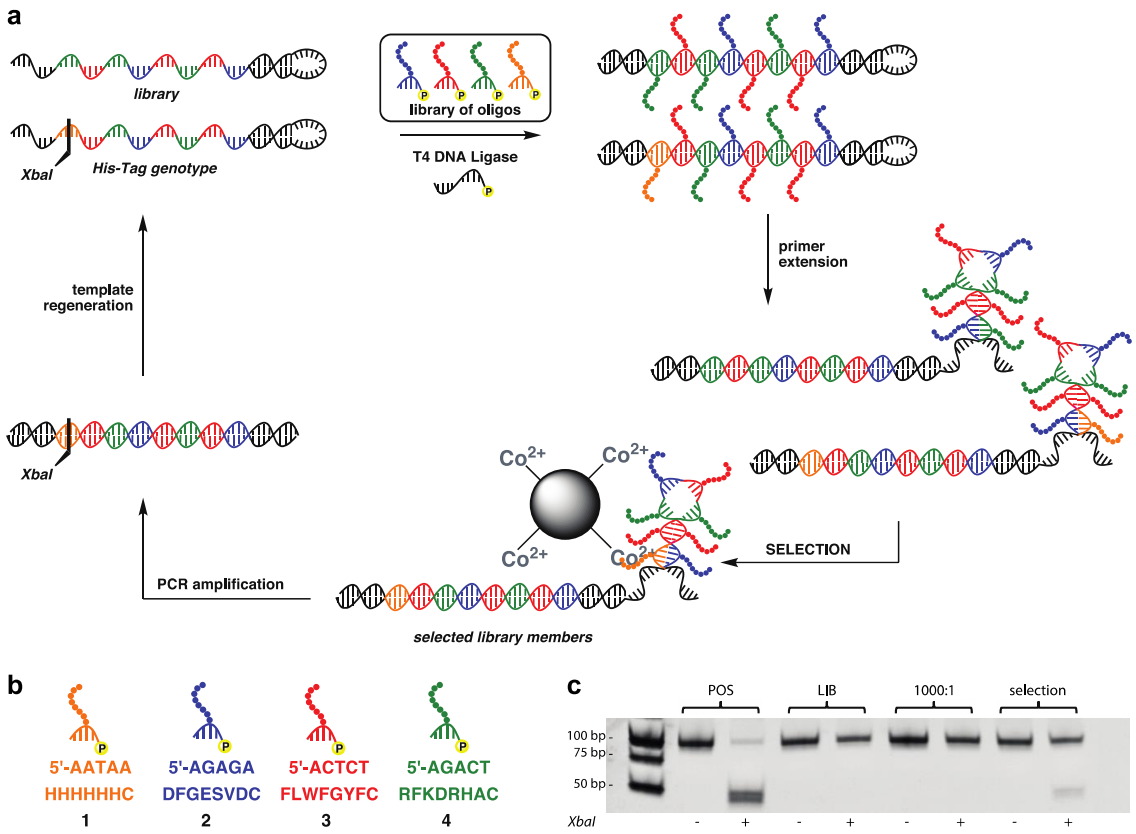


Figure 2.8: In vitro selection of DNA-scaffolded peptides for binding to Co^{2+} beads. (**a**) Complete cycle of in vitro selection of DNA-scaffolded peptides. His-tagged genotype (POS) is diluted 1000-fold into library (LIB). The library is then subjected to polymerization, primer extension-mediated strand displacement by *Bst* polymerase, selection against Co^{2+} magnetic particles, PCR amplification, and digestion by *XbaI* restriction enzyme. (**b**) Pentanucleotides used in the selection. Modification site is at the 5'-dA position. Conjugation of peptide is via a SMCC linker between the C6 amino dA nucleotide and the C-terminal cysteine residue of the peptide. (**c**) Non-denaturing PAGE analysis of *XbaI* digestion products after one round of selection, resulting in 190-fold enrichment of the genotype encoding the hexahistidine-tagged phenotype.

2.4 Conclusions and acknowledgements

In summary, we have developed a new approach for the generation of sequence-defined DNA-scaffolded peptides that uses T4 DNA ligase to catalyze the DNA-templated polymerization of peptide-modified 5'-phosphorylated pentanucleotides. Optimization of ATP concentration, template architecture, and peptide attachment to the pentanucleotide enabled efficient polymerization of these challenging substrates. Peptides ranging from 2-8 amino acids in length with a wide variety of functionality were found to be within the scope of this method. A four-codon library ranging from 0-40 % GC-content was used to demonstrate the high sequence specificity of the polymerization process. We anticipate expanding the codon set for this process to accommodate DNA scaffolding of a larger set of peptides. The ability to sequence-specifically incorporate multiple peptide fragments throughout an evolvable ssDNA polymer should enable the in vitro selection of ssDNA-scaffolded peptides that harness the power of heteromultivalency for molecular recognition of protein targets.

This work was supported by the NSF (DMR 1506667) and the Office for the Vice President of Research, University of Georgia. We thank the PAMS core facility at the University of Georgia for their help in the characterization of oligonucleotides. We would also like to thank Prof. Eric M. Ferreira and Prof. Vladimir V. Popik for helpful comments on the manuscript.

2.5 Experimental details

General information

Unless otherwise noted, all materials and compounds were prepared using commercially available reagents and used without further purification. Water was purified with Milli-Q purification system. DNA pentanucleotideswere synthesized on a BioautomationMermade 12 synthesizer. Peptides were purchased from Genscript. DNA oligonucleotides greater than five nucleotides in length were purchased from Integrated DNA Technologies. All materials and reagents used for oligonucleotide synthesis were purchased from GlenResearch. All oligonucleotides were synthesized and deprotected according to the manufacturer's protocols. Oligonucleotides were purified by reverse-phase high-pressure liquid chromatography (HPLC, Agilent 1260) using a C18 stationary phase (Eclipse-XDB C18, 5 μm,9.4 x 200 mm) and an acetonitrile/ 100 mMtriethylammonium acetate gradient. Oligonucleotide concentrations were quantitated by UV spectroscopy using a Nanodrop ND2000 spectrophotometer. Non-commercial oligonucleotides were characterized by LC/ESI-MS using a c18 column at ambient temperature with a mobile phase of 100% 6 mMtriethylammonium carbonate (TEAB) to 80% MeOH/20% TEAB over 10 minutes, and a flow rate of 0.3 mL/min; Oligonucleotides greater than 70 nucleotides in length were analyzed by PAGE.

DNA Sequences

The sequences below are written from 5'→3'. <P> = 5'Phosphate (Glen Research 10-1902); <Cam> = Amino-modifier C6 dC (Glen Research 10-1019); <Gam> = N2-Amino-modifier C6 dG (Glen Research 10-1529); <Aam> = Amino-modifer C6 dA (Glen Research 10-1089); <Tam> = Amino-modifer C6 dT (Glen Research 10-1039)

Templates

TH8B5c:/5Phos/TG CGA CGG CAG GTT CCC CTG CCG TCG CAA GAG TAG AGT AGA GTA GAG TAG AGT AGA GTA GAG TAG AGT C

TH8B5Pa: /5Phos/TG CGA CGG CAG GTT CCC CTG CCG TCG CAA GAG TAG AGT AGA GTA GAG TAG AGT AGA GTA GAG TAG AGT CAC GTG GAG CTC GGA TCC

TH8B5Pc: /5Phos/TG CGA CGG CAG GTT CCC CTG CCG TCG CAA GAG TAG AGT AGA GTA GAG TAG AGT AGA GTA GAG TAG AGT CAC GTG GAG CTC

T8B5H: AGA GTA GAG TAG AGT AGA GTA GAG TAG AGT AGA GTA GAG TAC GCT GCC GTC CCC TTG GAC GGC AGC GT

PaT8B5H: CAC GTG GAG CTC GGA TCC AGA GTA GAG TAG AGT AGA GTA GAG TAG AGT AGA GTA GAG TAC GCT GCC GTC CCC TTG GAC GGC AGC GT

PbT8B5H: CAC GTG GAG CTC GGA CCA AGA GTA GAG TAG AGT AGA GTA GAG TAG AGT AGA GTA GAG TAC GCT GCC GTC CCC TTG GAC GGC AGC GT

TH8C5c: /5Phos/TG CGA CGG CAG GTT CCC CTG CCG TCG CAA GAG AAG AGA AGA GAA GAG AAG AGA AGA GAA GAG AAG AGA C

TH8D5c: /5Phos/TG CGA CGG CAG GTT CCC CTG CCG TCG CAA GAG CAG AGC AGA GCA GAG CAG AGC AGA GCA GAG CAG AGC C

TH8E5c: /5Phos/TG CGA CGG CAG GTT CCC CTG CCG TCG CAA GAG GAG AGG AGA GGA GAG GAG AGG AGA GGA GAG GAG AGG C

TH8F5c: /5Phos/TG CGA CGG CAG GTT CCC CTG CCG TCG CAG GCG GGG CGG GGC GGG GCG GGG CGG GGC GGG GCG GGG CGG C

TH8G5c: /5Phos/TG CGA CGG CAG GTT CCC CTG CCG TCG CAT TAT TTT ATT TTA TTT TAT TTT ATT TTA TTT TAT TTT ATT C

TH8I5c： /5Phos/TG CGA CGG CAG GTT CCC CTG CCG TCG CAA GAT GAG ATG AGA TGA GAT GAG ATG AGA TGA GAT GAG ATG C

TH8J5c：/5Phos/TG CGA CGG CAG GTT CCC CTG CCG TCG CAA GTA GAG TAG AGT AGA GTA GAG TAG AGT AGA GTA GAG TAG C

TH8K5c： /5Phos/TG CGA CGG CAG GTT CCC CTG CCG TCG CAA TGA GAT GAG ATG AGA TGA GAT GAG ATG AGA TGA GAT GAG C

TH8L5c: /5Phos/TG CGA CGG CAG GTT CCC CTG CCG TCG CAT AGA GTA GAG TAG AGT AGA GTA GAG TAG AGT AGA GTA GAG C

TH4mixDc： /5Phos/TG CGA CGG CAG GTT CCC CTG CCG TCG CAT CTC TAG AGT TTA TTA GTC TC

T4mixCH： AGT CTT TAT TAG AGT TCT CTA CGC TGC CGT CCC CTT GGA CGG CAG CGT

SELMA-POS1: CTG TTG TTC CGC AGT CAC CTT TTA TTA GAG TTC TCT AGT CTT CTC TAG AGT AGT CTT CTC TCC CGT ACC CGT ATT TGG TGG CAA GGA TGA CAA GGA TTT TAT ATT TTA TAT TTT TAT TTT ATT ATC GGG TAC GGG

SELMA-NEG1: CTG TTG TTC CGC AGT CAC CTT AGA GTA GAG TAG AGT TCT CTT CTC TAG TCT AGT CTT CTC TCC CGT ACC CGT ATT TGG TGG CAA GGA TGA CAA GGA TTT TAT ATT TTA TAT TTT TAT TTT ATT ATC GGG TAC GGG

Primers

Pa(with TH8B5Pa): GGA TCC GAG CTC CAC GTG

Pc(with TH8B5Pc): GAG CTC CAC GTG

PPa(with PaT8B5H): /5Phos/GG ATC CGA GCT CCA CGT G

PPb(with PbT8B5H): /5Phos/TG GTC CGA GCT CCA CGT G

Poly primer1: /5phos/AA GGT GAC TGC GGA ACA ACA G

Aptfor: CCT TGT CAT CCT TGC CAC CA

Aptrev: CTG TTG TTC CGC AGT CAC CTT

Aptfor-bt: /5BiosG/CC TTG TCA TCC TTG CCA CCA

Amino-modified trinucleotides

PA5-1:<P><Aam>CTCT

PA5-2:<P>C<Aam>TCT

PA5-3:<P>CT<Aam>CT

PA5-4:<P>CTC<Aam>T

PA5-5:<P>CTCT<Aam>

PC5-1:<P><Cam>CTCT

PG5-1:<P><Gam>CTCT

PT5-1:<P><Tam>CTCT

PAs-1:<P><Aam>GAGA

PAs-2:<P><Aam>CTCT

PAs-3:<P><Aam>ATAA

PAs-4:<P><Aam>GACT

Terminator amino-modified trinucleotides

As-1: <Aam>GAGA

As-2:<Aam>CTCT

As-3: <Aam>ATAA

As-4:<Aam>GACT

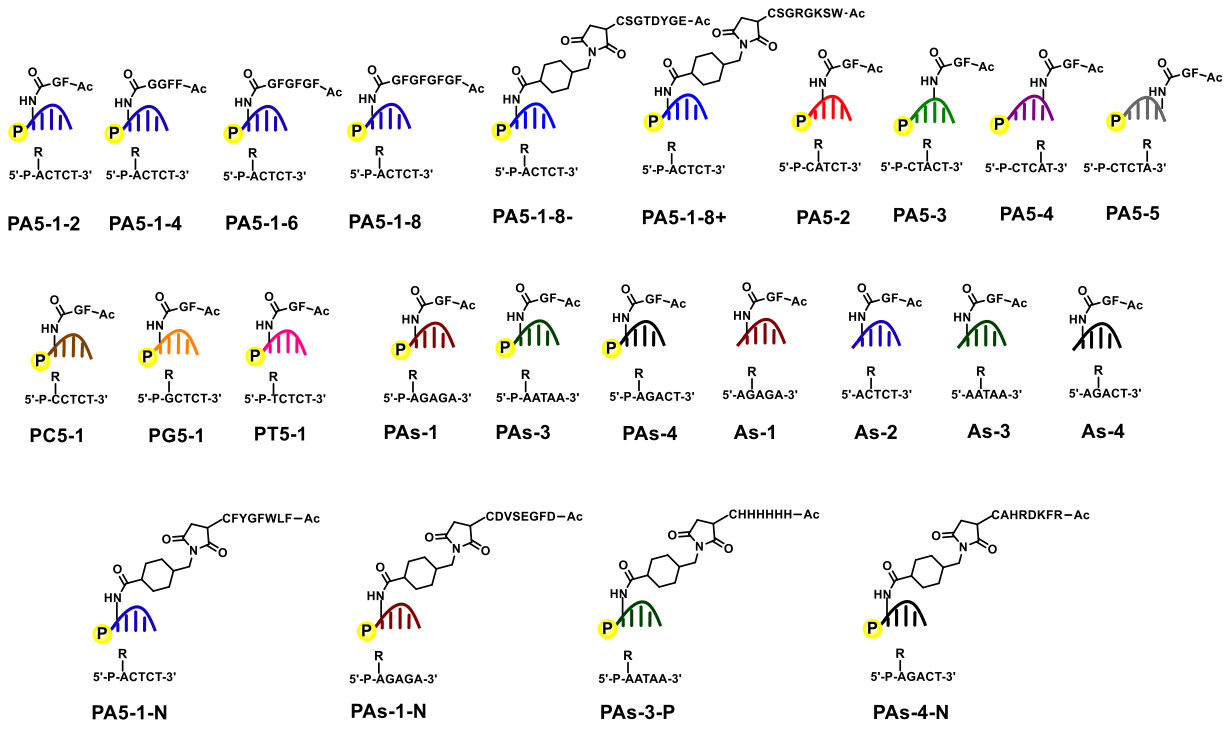


Figure S2.1: functionalized pentanucleotides libraries members

Synthesis of amino-modified pentanucleotides

Pentanucleotides were synthesized on a Mermaid 12 DNA synthesizer using a DMT-ON protocol on a 1 µmol scale (1000 Å CPG column). Amine-modifier C6 dA (Glen Research 10-1089), Amino-modifier C6 dC (Glen Research 10-1019), N2-Amino-modifier C6 dG (Glen Research 10-1529) and Amino-modifer C6 dT (Glen Research 10-1039), Chemical Phosphorylation Reagent II (10-1901) were incorporated as specified by the manufacturer.

Following synthesis, the oligonucleotide was cleaved from the resin by incubation at 65 °C in 500µL of a 1:1 mixture of ammonium hydroxide and methylamine for 15 minutes. The cleaved resin was filtered away by filtration, and the oligonucleotide was concentrated under reduced pressure using a speedvac. The residue was then taken up into 100 µL of H_2O, and purified using reverse-phase HPLC purification using a [10% acetonitrile in 0.1 M TEAA, pH 7] to [80% acetonitrile in 0.1 M TEAA, pH 7] solvent gradient with a column temperature of 45°C. The purified oligonucleotide was then incubated at room temperature in 500µL of 80% aqueous acetic acid for 1 h to cleave the DMT group, and then frozen and lyophilized. The oligonucleotide was incubated in 500µL 30% ammonium hydroxide at room temperature for 15 minutes to cleave the CPRII linker. Following deprotection, the oligonucleotide was concentrated under reduced pressure using a speedvac. The dried product was dissolved into 100 µL H_2O and subjected to reverse-phase HPLC purification using a [10% acetonitrile in 0.1 M TEAA, pH 7] to [80% acetonitrile in 0.1 M TEAA, pH 7] solvent gradient with a column temperature of 45°C. The purified oligonucleotide was dissolved in water.

Synthesis of pentanucleotides modified with FG-series peptides

To 215 µL of DMSO was added EDC (1200nmol in 12µL water), sNHS (3333nmol in 10µL 2/1 DMSO/water) and peptide (1250nmol in 12.5 µL of DMSO), and the reaction was incubated at room temperature for 45 minutes. After the initial incubation was added 5'-phosphorylated amino-modified pentanucleotide (25 nmol in 14.5 µL water) and 50µL of 500mM NEt_3-HCl (pH 10), and the reaction was incubated at room temperature overnight. The reaction was then quenched with 50µL of 500 mM Tris-HCl (pH 8) and incubated at room temperature for 1 h. The mixture was frozen and lyophilized to

dryness. The crude product was dissolved in 100µL water and then subjected to reverse-phase HPLC purification using a [10% acetonitrile in 0.1 M TEAA, pH 7] to [80% acetonitrile in 0.1 M TEAA, pH 7] solvent gradient with a column temperature of 45°C.

Functionalization of amino-modified pentanucleotides with Ac-EGYDTGSC-NH_2 and Ac-WSKGRGSC-NH_2

To 237µL of DMSO was added amino-modified pentanucleotides (25nmol in 14.5µL water), SMCC solution (1250nmol in 12.5µL DMSO) and 50µL of 500mM NEt_3-HCl (pH 10). The reaction was incubated at room temperature for 30 minutes. The reaction was then quenched with 50 µL of 500 mM Tris-HCl (pH 8) and incubated at room temperature for 1 h. The mixture was frozen and lyophilized to dryness. The crude product was dissolved in 100 µL water and then subjected to reverse-phase HPLC purification using a [10% acetonitrile in 0.1 M TEAA, pH 7] to [80% acetonitrile in 0.1 M TEAA, pH 7] solvent gradient with a column temperature of 45°C. The purified product was dissolved in 200µL 0.1M KH_2PO_4 (pH 7.2), and the peptide (100nmol in 200µL DMSO) was added to the solution, and the reaction was incubated at room temperature for 3 hours. The reaction mixture was frozen and lyophilized to dryness. The crude product was dissolved in 100 µL water and then subjected to reverse-phase HPLC purification using a [10% acetonitrile in 0.1 M TEAA, pH 7] to [80% acetonitrile in 0.1 M TEAA, pH 7] solvent gradient with a column temperature of 45°C.

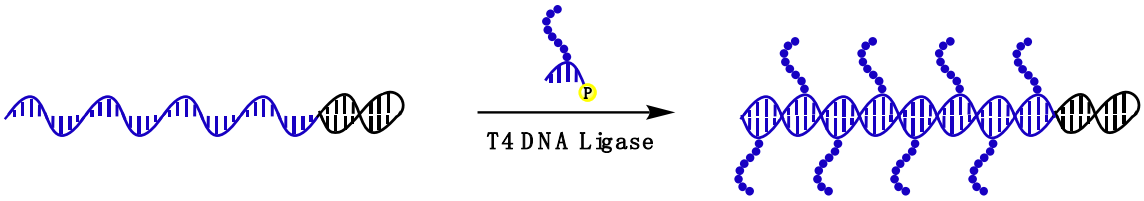
Mass spectrometric characterization of modified pentanucleotides

Pentanucleotide	**Sequence**	**Calculated mass**	**Observed mass**
PA5-1-2	<P>ACTCT	1877.5	1877.8
PA5-1-4	<P>ACTCT	2081.5	2081.5

PA5-1-6	<P>ACTCT	2285.7	2285.8
PA5-1-8	<P>ACTCT	2489.7	2489.1
PA5-1-8-	<P>ACTCT	2722.3	2722.1
PA5-1-8+	<P>ACTCT	2772.5	2772.3
PA5-2	<P>CATCT	1877.5	1877.5
PA5-3	<P>CTACT	1877.5	1877.4
PA5-4	<P>CTCAT	1877.5	1877.5
PA5-5	<P>CTCTA	1877.5	1877.6
PC5-1	<P>CCTCT	1907.5	1907.5
PG5-1	<P>GCTCT	1878.5	1878.5
PT5-1	<P>TCTCT	1908.5	1908.5
PAs-1	<P>AGAGA	1975.5	1975.6
PAs-3	<P>AATAA	1934.5	1934.6
PAs-4	<P>AGACT	1926.5	1926.5
As-1	AGAGA	1895.5	1895.8
As-2	ACTCT	1797.5	1797.7
As-3	AATAA	1854.5	1854.7
As-4	AGACT	1846.5	1846.6
PA5-1-N	<P>ACTCT	2972.8	2972.7
PAs-1-N	<P>AGAGA	2860.4	2860.0
PAs-3-P	<P>AATAA	2894.6	2894.0
PAs-4-N	<P>AGACT	2973.7	2973.7

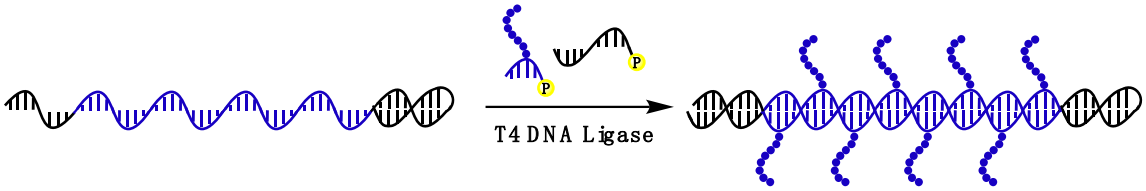
Table S2.1 Mass spectrometric characterization of modified pentanucleotides

Polymerization on a homo-octameric codon template



Scheme S2.1: Polymerization on a homo-octameric codon template (without primer)

In a PCR tube was added DNA template (15 pmol in 1.5 µL of water), 10µL of ligation buffer (132 mM Tris-HCl, 20 mM MgCl2, 2 mM dithiothreitol, 20% Polyethylene glycol (PEG 6000), pH 7.6), 5 µL of water and ATP (0.5nmol in 0.5 µL of water). The reaction mixture was heated to 94 °C for 2 minutes and then cooled down to 25 °C at the rate of 0.1 °C/s. In this PCR tube was then added functionalized pentanucleotides (480 pmol in 1 µL of water; 4 equivalents/codon), BSA (2 µg in 1 µL of water), 400 U of T4 DNA ligase (New England Biolabs, M0202L). The polymerization was performed at 25 °C for 24 h and then desalted by gel filtration using CENTRI • SEP Spin Columns (Princeton Separations) equilibrated with water. The crude polymerized material was separated for analysis using denaturing PAGE (15% TBE, 150 V, 55 °C) and then stained by ethidium bromide and imaged by UV illumination.

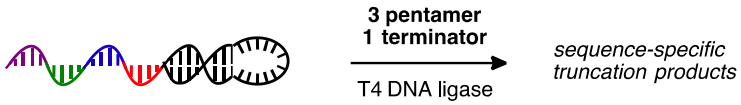


Scheme S2.2: Polymerization on a homo-octameric codon template (with primer)

In a PCR tube was added DNA template (15 pmol in 1.5 µL of water), primer (22.5 pmol in 2.25 µL of water) 10µL of ligation buffer (132 mM Tris-HCl, 20 mM MgCl2, 2 mM dithiothreitol, 20%Polyethylene glycol (PEG 6000), pH 7.6), 2.75 µL of water and ATP

(0.5nmol in 0.5 µL of water). The reaction mixture was heated to 94 °C for 2 minutes and then cooled down to 25 °C at the rate of 0.1 °C/s. In this PCR tube was then added functionalized pentanucleotides (480 pmol in 1 µL of water; 4 equivalents/codon), BSA (2 µg in 1 µL of water), 400 U of T4 DNA ligase (New England Biolabs, M0202L). The polymerization was performed at 25 °C for 24 h and then desalted by gel filtration using CENTRI • SEP Spin Columns (Princeton Separations) equilibrated with water. The crude polymerized material was separated for analysis using denaturing PAGE (15% TBE, 150 V, 55 °C) and then stained by ethidium bromide and imaged by UV illumination.

Sequence specificity assay by chain termination



Scheme S2.3: Terminator-mediated sequence specificity assay

In a PCR tube was added 15 pmol of template T4mixCH (AGT CTT TAT TAG AGT TCT CTA CGC TGC CGT CCC CTT GGA CGG CAG CGT), 10 µL of ligation buffer (132 mM Tris-HCl, 20 mM MgCl2, 2 mM dithiothreitol, 20% Polyethylene glycol (PEG 6000), pH 7.6), 2 µL water and ATP (0.5nmol in 0.5 µL water). The reaction mixture was heated to 94 °C for 2 minutes and then cooled down to 25 °C at the rate of 0.1 °C/s. In this PCR tube was added three of the four 5'-phosphorylated modified pentanucleotides (60pmol each from PAs-1, PAs-2(PA5-1), PAs-3 , PAs-4) with fourth modified pentanucleotide as a terminator (60 pmol from As-1, As-2, As-3 ,As-4), BSA (2 ug in 1 µL of water), and 400 U of T4 DNA ligase (New England Biolabs, M0202L) . The polymerization was performed at 25 °C for 24 h and then desalted by gel filtration using CENTRI • SEP Spin Columns (Princeton Separations) equilibrated with water. The crude polymerized material was separated for analysis using denaturing PAGE (15% TBE, 150

V，55°C) and then stained by ethidium bromide and imaged by UV illumination. In four separate experiments, the terminator was incorporated from the first codon to last codon.

In vitro mock selection of Co^{2+} binding for the enrichment of His-Tagged phenotypes

In a PCR tube was added premixed 1:1000 SELMA-POS1 to SELMA-NEG1 DNA template (15 pmol in 1.5 µL of water), poly primer1 (22.5 pmol in 2.25 µL of water) 10 µL of ligation buffer (132 mM Tris-HCl, 20 mM $MgCl_2$, 2 mM dithiothreitol, 20% Polyethylene glycol (PEG 6000), pH 7.6), 2.75 µL of water and ATP (0.5 nmol in 0.5 µL of water, 25 µM reaction concentration). The reaction mixture was heated to 94 °C for 2 minutes and then cooled down to 25 °C at the rate of 0.1 °C/s. The pentanucleotide library (120 pmol each), BSA(1 µL of 0.2mg/mL)and 400U of T4 DNA ligase was added to make a 20.25µL reaction volume. After 24h incubate under 25°C, the reaction mixture was purified with MinElute reaction clean up kit.

To the purified library in 10µL of water was added Aptfor primer (18.8 pmol), Isothermal amplification buffer (1X),dNTPs(200µM reaction concentration), 4U of Bst DNA polymerase(NEB, M0538L) and water up to 18.75µL. The mixture was added to a pre-heatedblock at 65°C for 5 minutes. The reaction mixture was purified with MinElute reaction clean up kit.

The DNA-displayed DNA-scaffolded library was taken up into200µL of 1X binding buffer (50 mM Sodium Phosphate, pH 8.0, 300 mM NaCl, 0.01% Tween-20). 2µL of Co^{2+} magnetic beads (Dynabeads His-Tag Isolation and Pulldown, Invitrogen, 10103D)was added to the library sample and incubated on a rotary for 30 minutes at 25°C. The magnetic beads were isolated by use of a magnet, and the supernatant was discarded. The beads were then washed six times with 1X binding buffer. After removing

the binding buffer, add 100µL His elution buffer (300 mM Imidazole, 50 mM sodium phosphate, 300 mM NaCl, 0.01% Tween-20, pH 8.0) then incubate on rotary for 5 minutes to elute the hexahistidine-tagged library members. The solution was purified with a CENTRI•SEP Spin Column (Princeton Separations) to remove the excess salt and imidazole prior to PCR amplification.

The eluted library members were subjected to PCR amplification. Thus, to the surviving genotypes in 10 µL of water was added the Aptfor-bt and Aptrev primers (60 pmol each), thermopol buffer (1X), dNTPs (200µM reaction volume) and 4U Vent(exo-) polymerase (NEB, M0257L) and water up to200µL volume. The PCR cycles were as follows: 95 °C for 90 s to denature, then 95 °C for 15 s, 64 °C for 20 s, 72 °C for 10 s for 20 cycles. Following the completion of 20 cycles, 15U of Exo I and Exonuclease I (NEB, M0293L), Reaction Buffer (1X) was added to the amplification mixture and allowed to incubate at 37°C for 30 minutes. The digest was subsequently purified using a MinElute reaction clean up kit.

To assess the enrichment over the round, a digestion was performed using a restriction enzyme that specifically targeted the POS genotype. Thus, the amplified DNA library (1 pmol), was taken up into Cut Smart Buffer (1X), and 2U of *XbaI* (NEB, R0145L) and water was added up to 10 µL. The mixture was incubated at 37°C for 20 hours. The digested product was then purified with MinElute reaction clean up kit and the digestion products were analyzed by 15% TBE non-denaturing PAGE.

2.6 Supporting Results

Influence of modified nucleobase identity on polymerization

In a PCR tube was added DNA template (15 pmol in 1.5 µL of water), 10µL of ligation buffer (132 mM Tris-HCl, 20 mM MgCl2, 2 mM dithiothreitol, 20% Polyethylene glycol (PEG 6000), pH 7.6), 5 µL of water and ATP (0.5 nmol in 0.5 µL of water). Pre-heated the mixture to 94°C in 2 minutes and cooled down to 25°C at the rate of 0.1°C /s and keep the temperature for 30 seconds. In this PCR tube was then added functionalized pentanucleotides (480 pmol in 1 µL of water; 4 equivalents/codon), BSA (2 ug in 1 uL of water), 400 U of T4 DNA ligase (New England Biolabs, M0202L). The polymerization was performed at 25 °C for 24 h and then desalted by gel filtration using CENTRI • SEP Spin Columns (Princeton Separations) equilibrated with water. The crude polymerized material was separated for analysis using denaturing PAGE (15% TBE, 150 V, 55 °C) and then stained by ethidium bromide and imaged by UV illumination. Modifications at dT, dC, and dG failed to yield full-length product**.**

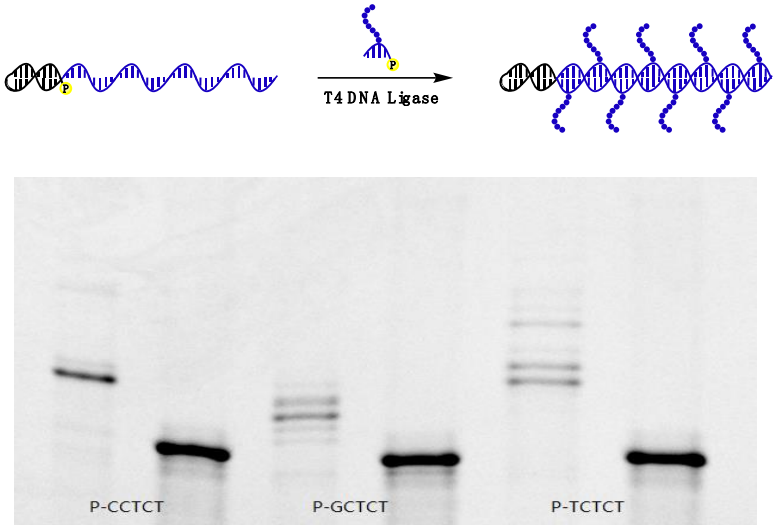


Figure S2.2 Influence of nucleobase identity at position one when modified with Ac-Phe-Gly.

2.7 Reference

1. Mammen, M.; Choi, S.-K.; Whitesides, G. M., Polyvalent Interactions in Biological Systems: Implications for Design and Use of Multivalent Ligands and Inhibitors. *Angew. Chem., Int. Ed. Engl.* **1998,** *37* (20), 2754-2794.

2. Mahon, C. S.; Fulton, D. A., Mimicking nature with synthetic macromolecules capable of recognition. *Nat Chem* **2014,** *6* (8), 665-72.

3. Fasting, C.; Schalley, C. A.; Weber, M.; Seitz, O.; Hecht, S.; Koksch, B.; Dernedde, J.; Graf, C.; Knapp, E. W.; Haag, R., Multivalency as a chemical organization and action principle. *Angew Chem Int Ed Engl* **2012,** *51* (42), 10472-98.

4. Kiessling, L. L.; Gestwicki, J. E.; Strong, L. E., Synthetic multivalent ligands as probes of signal transduction. *Angew Chem Int Ed Engl* **2006,** *45* (15), 2348-68.

5. Fu, J.; Liu, M.; Liu, Y.; Yan, H., Spatially-interactive biomolecular networks organized by nucleic acid nanostructures. *Acc Chem Res* **2012,** *45* (8), 1215-26.

6. Wilner, O. I.; Willner, I., Functionalized DNA nanostructures. *Chem Rev* **2012,** *112* (4), 2528-56.

7. Diezmann, F.; Seitz, O., DNA-guided display of proteins and protein ligands for the interrogation of biology. *Chem Soc Rev* **2011,** *40* (12), 5789-801.

8. McLaughlin, C. K.; Hamblin, G. D.; Sleiman, H. F., Supramolecular DNA assembly. *Chem Soc Rev* **2011,** *40* (12), 5647-56.

9. Temme, J. S.; MacPherson, I. S.; DeCourcey, J. F.; Krauss, I. J., High temperature SELMA: evolution of DNA-supported oligomannose clusters which are tightly recognized by HIV bnAb 2G12. *J Am Chem Soc* **2014,** *136* (5), 1726-9.

10. Kazane, S. A.; Axup, J. Y.; Kim, C. H.; Ciobanu, M.; Wold, E. D.; Barluenga, S.; Hutchins, B. A.; Schultz, P. G.; Winssinger, N.; Smider, V. V., Self-assembled antibody multimers through peptide nucleic acid conjugation. *J Am Chem Soc* **2013,** *135* (1), 340-6.

11. MacPherson, I. S.; Temme, J. S.; Habeshian, S.; Felczak, K.; Pankiewicz, K.; Hedstrom, L.; Krauss, I. J., Multivalent glycocluster design through directed evolution. *Angew Chem Int Ed Engl* **2011,** *50* (47), 11238-42.

12. Ghosh, P. S.; Hamilton, A. D., Noncovalent template-assisted mimicry of multiloop protein surfaces: assembling discontinuous and functional domains. *J Am Chem Soc* **2012,** *134* (32), 13208-11.

13. Williams, B. A.; Diehnelt, C. W.; Belcher, P.; Greving, M.; Woodbury, N. W.; Johnston, S. A.; Chaput, J. C., Creating protein affinity reagents by combining peptide ligands on synthetic DNA scaffolds. *J Am Chem Soc* **2009,** *131* (47), 17233-41.

14. Gorska, K.; Huang, K.-T.; Chaloin, O.; Winssinger, N., DNA-Templated Homo- and Heterodimerization of Peptide Nucleic Acid Encoded Oligosaccharides that Mimick the Carbohydrate Epitope of HIV. *Angew. Chem., Int. Ed. Engl.* **2009,** *48* (41), 7695-7700.

15. Yin, H.; Hamilton, A. D., Strategies for targeting protein-protein interactions with synthetic agents. *Angew Chem Int Ed Engl* **2005,** *44* (27), 4130-63.

16. Moreira, I. S.; Fernandes, P. A.; Ramos, M. J., Hot spots--a review of the protein-protein interface determinant amino-acid residues. *Proteins* **2007,** *68* (4), 803-12.

17. Hili, R.; Niu, J.; Liu, D. R., DNA ligase-mediated translation of DNA into densely functionalized nucleic acid polymers. *J Am Chem Soc* **2013,** *135* (1), 98-101.

18. Kotler, L. E.; Zevin-Sonkin, D.; Sobolev, I. A.; Beskin, A. D.; Ulanovsky, L. E., DNA sequencing: modular primers assembled from a library of hexamers or pentamers. *Proc Natl Acad Sci U S A* **1993,** *90* (9), 4241-5.

19. Hermanson, G. T., Bioconjugate techniques. Third edition. ed.; Academic Press is an imprint of Elsevier: London, 2013.

20. Zimmerman, S. B.; Minton, A. P., Macromolecular crowding: biochemical, biophysical, and physiological consequences. *Annu Rev Biophys Biomol Struct* **1993,** *22*, 27-65.

21. Cherepanov, A. V.; de Vries, S., Kinetics and thermodynamics of nick sealing by T4 DNA ligase. *Eur J Biochem* **2003,** *270* (21), 4315-25.

22. Yu, H.; Zhang, S.; Chaput, J. C., Darwinian evolution of an alternative genetic system provides support for TNA as an RNA progenitor. *Nat Chem* **2012,** *4* (3), 183-7.

CHAPTER 3

Fidelity of the DNA Ligase-Catalyzed Scaffolding of Peptide Fragments on Nucleic Acid Polymers[1]

1. Guo, C.; Hili, R. *Bioconjug Chem* **2017,** *28* (2), 314-318. Reproduced with the permission from Guo, C.; Hili, R. *Bioconjug Chem* **2017,** *28* (2), 314-318. Copyright 2016 American Chemical Society.

3.1 Abstract

We describe the development and analysis of the T4 DNA ligase-catalyzed DNA templated polymerization of pentanucleotides modified with peptide fragments toward the generation of ssDNA-scaffolded peptides. A high-throughput duplex DNA sequencing method was developed to facilitate the determination of fidelity for various codons sets and library sizes used during the polymerization process. With this process, we identified several codon sets that enable the efficient and sequence-specific incorporation of peptide fragments along a ssDNA template at fidelities up to 99% and with low sequence bias. These findings mark a significant advance in generating evolvable biomimetic polymers and should find ready application to the in vitro selection of molecular recognition.

3.2 Introduction

Protein-protein interactions (PPIs) are central to biological function. These interactions give rise to a diverse set of processes ranging from DNA replication, transcription,[1] protein modification[2] and signal transduction[3]. Protein-protein interfaces, which drive these interactions, are generally large surfaces with shallow topologies spanning from 1,500—3,000 $Å^2$.[4] At the interface, heteromultivalent interactions in the form of peptide fragments forge complementary regions of hydrophobicity and electrostatic forces to enable highly specific assembly. As with most protein function, aberrant PPIs have been implicated in a broad scope of human disease.[5] Thus, molecules that achieve the selective interrogation of PPIs have garnered significant interest as possible therapeutics.[6] Despite recent advances in targeting PPIs using small-molecules, the discovery of such molecules

remain a challenging pursuit due to the large surface areas of protein-protein interfaces.[7] One particularly attractive approach to disrupt these pathological interaction is the rationally design[8] or evolution[9] of biopolymers that mimic the interfacial surface of one of the protein partners. This strategy has been central to the development of new molecular therapeutics that disrupt PPIs interface.[6]

Single-stranded nucleic acid polymers decorated in a sequence- defined manner with peptide fragments are promising candidates to mimic the surface of proteins and serve as a new class of high-affinity reagents. Indeed, there has been significant interest in the sequence-defined display of peptide fragments and proteins on nucleic acid scaffolds.[10] While advances have been made in this area, most of these approaches implement a rigid duplex nucleic acid scaffold for the multivalent display of the peptide fragments, and thus do not take full advantage of the scope of molecular architectures available to ssDNA. Furthermore, most current approaches do not accommodate the incorporation and heteromultivalent display of several different peptide fragments in a library format, and thus preclude the evolution of the nucleic acid scaffold for optimal display of multiple ligands.

To this end, we recently developed the Ligase-catalyzed OligOnucleotide PolymERization (LOOPER) method, and demonstrated its ability in the sequence-defined incorporation of peptide fragments along a ssDNA template.[11] The method hinges upon the T4 DNA ligase-catalyzed DNA-templated polymerization of 5'-phosphorylated pentanucleotides that are modified with peptide fragments (**Figure 3.1**). While promising as a general method to evolve protein surface mimetics, the codon set of the system, and thus the number of unique peptide fragments that could be incorporated, has hitherto been

limited to four unique sequences. Therefore, we sought to expand the codon set of this process beyond four sequences, while also developing a high-throughput method to assess the fidelity of incorporation in a library context.

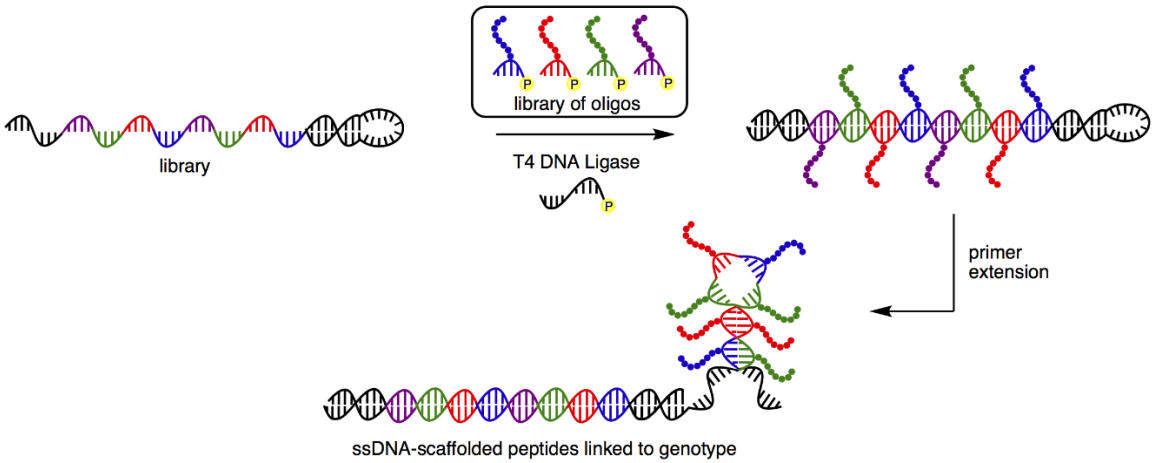


Figure 3.1: Generation of DNA-scaffolded peptides using LOOPER. Peptides of interest are covalently linked to the C6 amino dA on the pentanucleotides through corresponding reactive linker.

3.3 Results and discussions

The original codon set that was used to incorporate peptide fragments using LOOPER comprised TTATT, TCTCT, AGAGT, AGTCT. While these codons enabled efficient and highly specific copolymerization, the set was not ideal for the ready synthesis of combinatorial template libraries using commercially available mixed bases, which greatly limited its use for in vitro evolution of protein surface mimetics. Accordingly, we explored alternative codon sets that would satisfy the following conditions: (*i*) high-fidelity and high efficiency polymerization; (*ii*) in-frame annealing and polymerization; (*iii*) ready combinatorial library synthesis using mixed bases; (*iv*) larger codon sets; and (*v*) accessible generation of a library of well-folded ssDNA scaffolds. Using high-

throughput DNA duplex sequencing data garnered from LOOPER products generated using small-molecule modified pentanucleotides,[12] we selected four larger codon sets that demonstrated high fidelity and low sequence bias. We determined that the codon sets WSWST, SWSWT, SNWWT, and SNNWT (where S = C or G; W = A or T; and N = A, C, G, or T) each satisfied these requirements. Importantly, the corresponding 40 nucleotide libraries generated as reverse complements from these codon sets demonstrated comparable or better calculated[13] average folding thermodynamics to an N_{40} library (**Figure S3.5**). This suggests that these libraries will allow access to well-folded ssDNA scaffolds for peptide display, which will likely be a critical feature during selections.

Due to presence of peptide fragments on a ssDNA scaffold, the thermodynamics of folding might be different from its native form. This could significantly impact the potential for the in vitro selection of well-defined protein surface mimetics using this scaffolding strategy. To experimentally examine the influence of peptide fragments on the thermodynamics of folding of a ssDNA scaffold, we compared the thermal melting (T_m) of a specific 64 nt ssDNA both with and without conjugated peptide fragments (**Figure 3.2**). The peptide-modified ssDNA was adorned with eight octapeptide fragments (GSASIFLY) and was assembled in a preparative synthesis using the LOOPER method. Thermal melting was determined at 0.2 µM ssDNA concentration in a buffer comprising 0.15 M NaCl and 0.015M sodium citrate at pH 7. The T_m for the unmodified ssDNA was 66 °C, while the peptide-modified ssDNA melted at 74 °C. While this increase in T_m is likely due to stabilizing peptide-peptide or peptide-DNA

interactions, the well-behaved thermal melting suggests that multiple peptide fragments will be accommodated on folded ssDNA scaffolds.

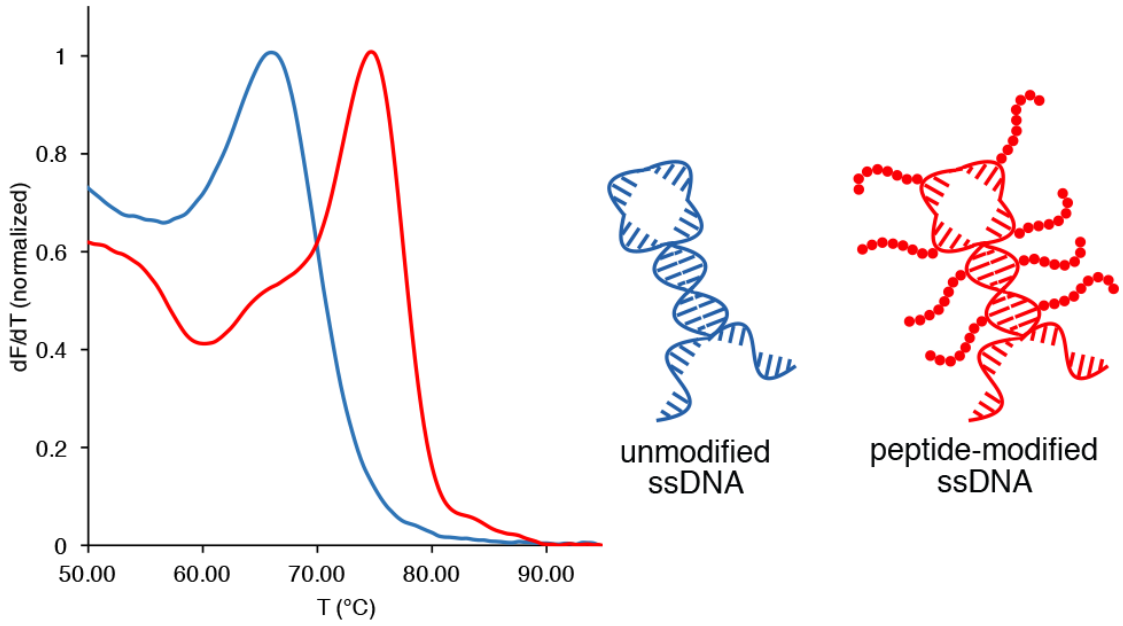


Figure 3.2: Thermal melting comparison between an unmodified 64 nt ssDNA, and the same sequence generated using LOOPER with pentanucleotides modified with GSASIFLY. Thermal melting profiles were generating by SYBR green melt assay.

We sought to apply the duplex DNA sequencing method[12,14] to assess the fidelity of LOOPER with libraries of peptide-modified pentanucleotides. As the ssDNA-scaffolded peptide products were not amenable to PCR amplification due to interference by the peptide fragments, we designed the pentanucleotide co-monomers to contain a cleavable linker between the peptide and the pentanucleotide. This would allow cleavage of the peptide fragment prior to PCR amplification and sequencing. Our initial approach was to use the bis(2-(succinimidooxycarbonyloxy)ethyl) sulfone (BSOCOES) cleavable linker to attach the amine-modified pentanucleotide libraries to the N-terminus of peptides; however, this linker was partially cleaved during the LOOPER process, which would

have resulted in skewing of fidelity levels. Thus, we used the disuccinimidyl tartrate (DST) cleavable linker to conjugate peptides to the oligo libraries (**Figure 3.3a**). We envisaged that the use of DST would enable the ready DNA sequencing of the duplex products following periodate-mediated linker cleavage to remove the large peptide fragments. The oxidative cleavage would result in an aldehyde scar; however, similarly sized modifications have been previously tolerated by duplex DNA sequencing.[12,14]

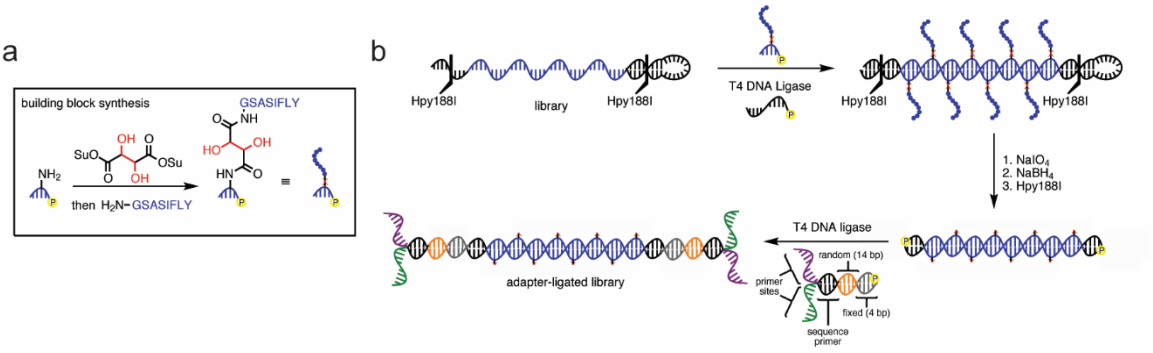


Figure 3.3: Duplex sequencing sample preparation. (**a**) Conjugation of a model peptide fragment to amino-modified pentanucleotide via DST linker to generate cleavable building blocks for LOOPER. Su = succinimidyl. (**b**) Optimized strategy for high-throughput duplex DNA sequencing of peptide-modified DNA generated by LOOPER. Note that the restriction enzyme Hpy188I was used for the WSWST codon set, while HpyCH4III was used for SWSWT, SNWWT, and SNNWT codon sets.

Our first attempts to polymerize the DST linked peptide-modified pentanucleotides using a two-primer system, as required for the duplex DNA sequencing method, failed to yield full-length product. Since the original LOOPER process for peptide-modified pentanucleotides was achieved on a hairpin architecture, we reasoned that the increased stability of the hairpin structure facilitated the polymerization. While it was conceivable that a longer primer could mitigate this issue, we were concerned about undesired

mispriming that might occur in a large library format, which would convolute sequencing data. Thus, we explored the possibility of using the original hairpin architecture for the polymerization process, followed by restriction digestion to permit ligation of the sequencing adapter. We chose HpyCH4III to perform the restriction digest for SWSWT, SNWWT and SNNWT codon sets, as its nucleotide-recognition. (**Figure S3.4**)

We speculated that the aldehyde function group resulting from oxidative cleavage was inhibiting T4 DNA ligase function. Thus, we reduced the aldehyde scars with $NaBH_4$ and proceeded with the adapter ligation. Gratifyingly, removal of the aldehydes enabled efficient bis-ligation. The optimized approach is summarized in **Figure 3.3b**. To the best of our knowledge, this is the first observation of proximal aldehyde inhibition of T4 DNA ligase activity; indeed, the presence of an aldehyde functional group on pentanucleotides during the LOOPER process greatly decreases the yield of the full-length polymer product (**Figure S3.2**). It is probable that the active site lysine nucleophile (Lys_{159}) of T4 DNA ligase may form a stable imine or a hemiaminal with the aldehyde-modified DNA, thus inhibiting the ligation process. As an alternative strategy, we also attempted to cleave the peptide fragments after adapter ligation to circumvent the aldehyde intermediate during the adapter ligation step. However, in our hands this resulted in somewhat lower yields of the bis-ligated product.

Using the optimized duplex DNA sequencing approach, we assessed the fidelity of LOOPER when polymerizing peptide-modified pentanucleotides along different codon sets. We chose GSASIFLY as the model peptide for conjugation to the pentanucleotide co-monomer library due to its ease of conjugation and representation of different hydrophobic residues. This would allow us to assess the LOOPER process across each

codon within the set. Four separate template libraries were designed with two primer binding sites flanking a reading frame comprising eight consecutive repeats of a specific codon set and the sequencing results from each library are summarized in **Table 1**. It is important to note that we were unable to examine LOOPER with codon sets that were greater than 64 members, as the yield of the full-length product dropped markedly at this level of complexity (see **Figure S3.3**). Following sequencing analysis, we observed excellent fidelities with all codon sets despite the potential for mis-incorporation of competing single-nucleotide mismatched co-monomers. As anticipated, the smaller library sizes resulted in higher fidelities and yield of full-length products. We also analysed the codon bias of LOOPER within each codon set. To do this we calculated the frequency of each codon in the template library and compared it with the frequency of each codon in the product library (**Figure 3.4**). Codon bias increased significantly as library size increased. As a metric for codon set bias, we determined the standard deviation of enrichment across each codon set, where enrichment is the frequency of a specific codon in the product divided by frequency of the same codon in the template (**Table 3.1**).

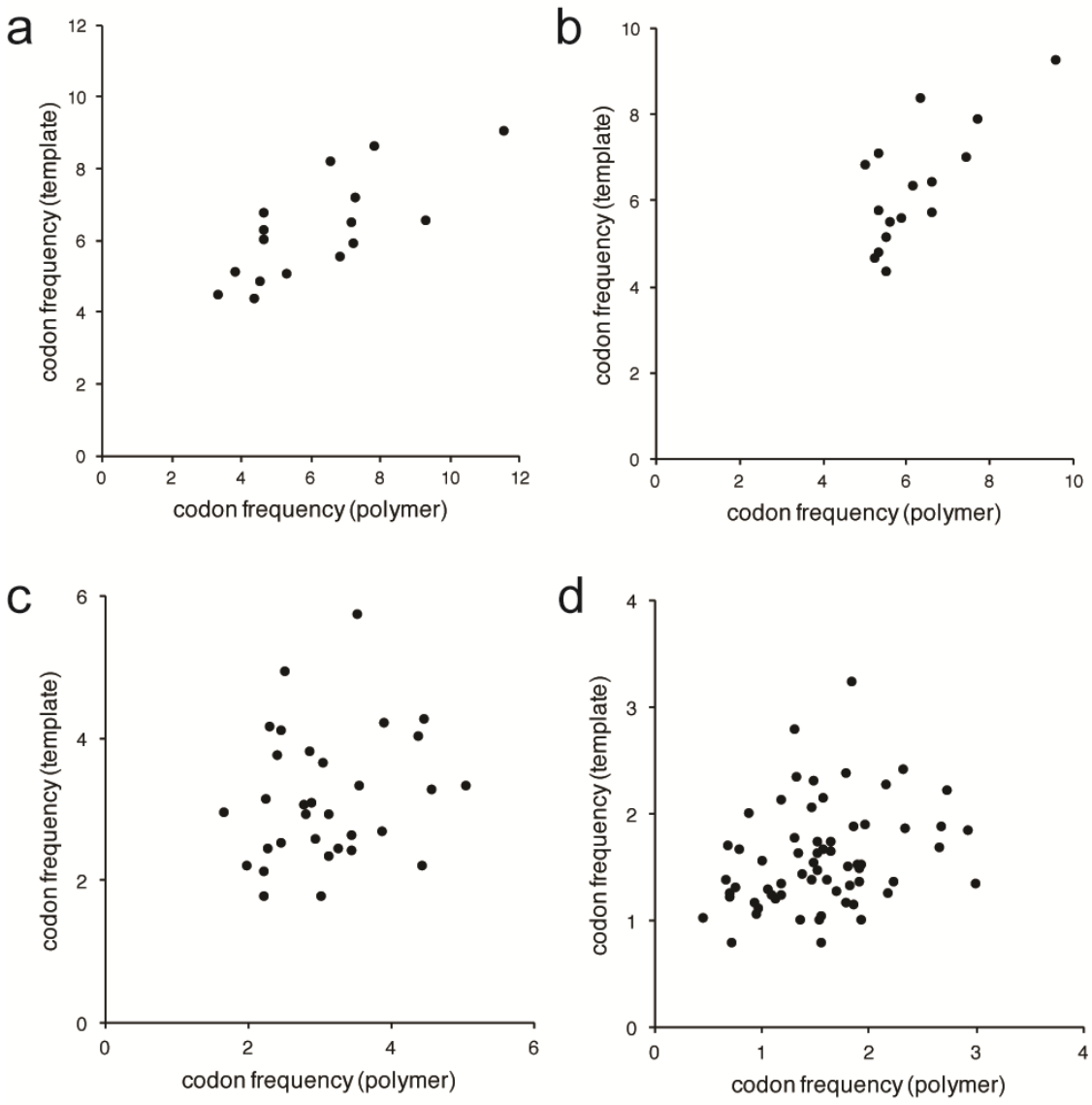


Figure 3.4: Codon bias observed for the various codon sets during LOOPER with corresponding peptide-modified pentanucleotide libraries. (**a**) WSWST; (**b**) SWSWT; (**c**) SNWWT; (**d**) SNNWT.

Entry	Pentanucleotide[a]	Codon set	Set size	Yield	Reads[b]	Bias[c]	Fidelity[d]
1	5'P-A*SWSW	WSWST	16	64%	141208	0.22	99.00%
2	5'P-A*WSWS	SWSWT	16	47%	168372	0.15	98.20%
3	5'P-A*WWNS	SNWWT	32	25%	110528	0.36	98.00%
4	5'P-A*WNNS	SNNWT	64	20%	41325	0.42	94.90%

Table 3.1: Fidelity of LOOPER with peptide-conjugated pentanucleotides. [a] pentanucleotides were modified with GSASIFLY at 5'dA, denoted by an asterisk. [b] Number of codon reads from sequencing data. [c] Bias was determined by the standard deviation of frequency of codons in polymer divided by the frequency of the codons in the original template. [d] Fidelity was calculated for pentanucleotide incorporation; the fidelity is considerably higher if evaluated at the single nucleotide level.

Due to the superior performance of the 16-membered codon set WSWST with respect to polymerization yield, fidelity, and codon bias, we believe that this set represents the best system for the in vitro selection of ssDNA-scaffolded peptides for molecular function. We thus examined the fidelity of the codon set at the single codon level (**Table 3.2**). While all codons performed with excellent fidelity during LOOPER, there was a 9-fold increase in error rate between the highest and lowest fidelity codon (TCACT and ACTCT, respectively); similar ranges in fidelity were observed with the other codon sets examined in this study (**Table S4.1**). A priori, it is difficult to speculate how this range of

fidelities within the codon set will affect codon distribution throughout iterative rounds of in vitro selection.

Entry	Pentanucleotide[a]	Codon	Enrich[b]	Fidelity[c]
1	5’P-A*GTGA	TCACT	1.25	99.6%
2	5’P-A*GTCA	TGACT	1.23	99.5%
3	5’P-A*CTGT	ACAGT	1.07	99.5%
4	5’P-A*CTGA	TCAGT	1.12	99.4%
5	5’P-A*CAGT	ACTGT	0.79	99.3%
6	5’P-A*CACT	AGTGT	1.03	99.2%
7	5’P-A*GACA	TGTCT	0.91	99.1%
8	5’P-A*CACA	TGTGT	1.28	99.1%
9	5’P-A*CTCA	TGAGT	1.43	99.0%
10	5’P-A*GAGA	TCTCT	0.70	99.0%
11	5’P-A*CAGA	TCTGT	0.81	98.9%
12	5’P-A*CTCT	AGAGT	0.95	98.9%
13	5’P-A*GACT	AGTCT	0.75	98.5%
14	5’P-A*GTCT	AGACT	0.77	98.4%
15	5’P-A*GTGT	ACACT	1.03	97.7%
16	5’P-A*GAGT	ACTCT	0.77	96.6%

Table 3.2: Fidelity of individual pentanucleotides during LOOPER

[a] pentanucleotides were modified with GSASIFLY at 5’dA, denoted by an asterisk. [b] Enrichment was calculated as (freq. polymer/freq. template). [c] Fidelity was calculated for

pentanucleotide incorporation; the fidelity is considerably higher if evaluated at the single nucleotide level.

3.4 Conclusions and acknowledgements

In summary, our findings reported herein demonstrate that the LOOPER process can combinatorially generate ssDNA-scaffolded peptide libraries from co-monomer libraries comprising up to 64 members. The 16-membered codon sets, WSWST and SWSWT, were particularly effective at generating DNA-scaffolded peptide libraries with good efficiencies and at fidelities of up to 99 %. This approach should enable the incorporation of 16 different peptides throughout an evolvable ssDNA polymer, with the potential to apply this technology to other types of fragments, including carbohydrates or synthetic oligomers. The ability to sequence-specifically incorporate multiple peptide fragments throughout an evolvable ssDNA polymer should enable the in vitro selection of ssDNA-scaffolded peptides that harness the power of heteromultivalency for molecular recognition of protein targets.

This work was supported by the NSF (DMR 1506667) and the Office for the Vice President of Research, University of Georgia. We would like to thank Dr. Saravanaraj Ayyampalayam and Yi Lei for help with the analysis of DNA sequencing data and the PAMS core facility at the University of Georgia for their help in the characterization of oligonucleotides.

3.5 Experimental details

General information

Unless otherwise noted, water was purified with ELGA Flex 3 purification system. DNA oligonucleotides without amine modification were purchased from Integrated DNA Technologies. DNA oligonucleotides with amine modification were synthesized on a Bioautomation Mermade 12 synthesizer. All materials and reagents used for oligonucleotide synthesis were purchased from Glen Research. All oligonucleotides were synthesized and deprotected according to the manufacturer's protocols. Oligonucleotides were purified by reverse-phase high-pressure liquid chromatography (HPLC, Agilent 1260) using a C18 stationary phase (Eclipse-XDB C18, 5 μm, 9.4 x 200 mm) and an acetonitrile/100 mM triethylammonium acetate gradient. Oligonucleotide concentrations were quantitated by UV spectroscopy using a Nanodrop ND2000 spectrophotometer.

DNA sequences

The DNA sequences below are written from 5'→3'. The peptide sequences are written from N term→C term. <Aam> = Amino-modifer C6 dA; <N>=A/T/C/G; <S>=G/C; <W>=A/T

Templates

MeltT template /5BiosG/ GAT TCG CCT GCC GTC GCA ACA GTA CAG TAC AGT ACA GTA CAG TAC AGT ACA GTA CAG TAC GCT CTG AGC CCC TTG GCT CAG AGC GT

MeltT natural strand GAG CGT ACT GTA CTG TAC TGT ACT GTA CTG TAC TGT ACT GTA CTG TTG CGA CGG CAG GCG AAT C

PASWSW template GAT TCG TCA GAC GTC GCA WSW STW SWS TWS WST WSW STW SWS TWS WST WSW STW SWS TAC GCT CTG AGC CCC TTG GCT CAG AGC GT

PAWSWS template GAT TCC ACA GTC GTC GCT SWS WTS WSW TSW SWT SWS WTS WSW TSW SWT SWS WTS WSW TAC GCA CTG TGC CCC TTG GCA CAG TGC GT

PAWWNS template GAT TCC ACA GTC GTC GCT SNW WTS NWW TSN WWT SNW WTS NWW TSN WWT SNW WTS NWW TAC GCA CTG TGC CCC TTG GCA CAG TGC GT

PAWNNS template GAT TCC ACA GTC GTC GCT SNN WTS NNW TSN NWT SNN WTS NNW TSN NWT SNN WTS NNW TAC GCA CTG TGC CCC TTG GCA CAG TGC GT

Primers

MeltT primer /5Phos/TG CGA CGG CAG GCG AAT C

5'-primer Hpy188I /5phos/TGC GAC GTC TGA CGA ATC

5'primer HpyCH4III /5phos/AGC GAC GAC TGT GGA ATC

Adapter A AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC T

Adapter B /5phos/ACT GNN NNN NNN NNN NNN AGA TCG GAA GAG CAC ACG TCT GAA CTC CAG TCA C

iTruS_i7_D701 CAA GCA GAA GAC GGC ATA CGA GAT ATT ACT CGG TGA CTG GAG TTC AG

iTruS_i7_D702 CAA GCA GAA GAC GGC ATA CGA GAT TCC GGA GAG TGA CTG GAG TTC AG

iTruS_i7_D703 CAA GCA GAA GAC GGC ATA CGA GAT CGC TCA TTG TGA CTG GAG TTC AG

iTruS_i7_D704 CAA GCA GAA GAC GGC ATA CGA GAT GAG ATT CCG TGA CTG GAG TTC AG

Pentanucleotides

PA-16-1 /5Phos/<Ama>SWSW

PA-16-2 /5Phos/<Ama>WSWS

PA-32 /5Phos/<Ama>WWNS

PA-64 /5Phos/<Ama>WNNS

PMeltT /5Phos/<Ama>CTGT

Peptides

GSASIFLY-NH_2

Synthesis of functionalized pentanucleotides

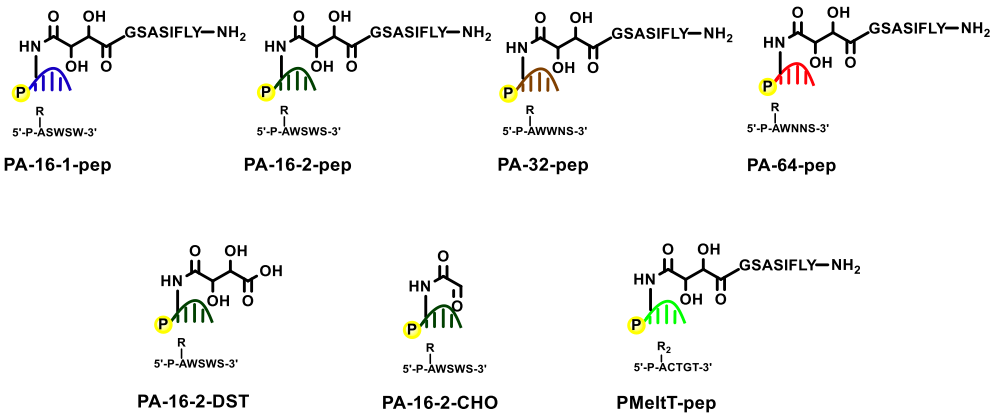


Figure S3.1: Modified pentanucleotides applied in this paper

Synthesis of amino-modified pentanucleotides

Pentanucleotides were synthesized on a Mermaid 12 DNA synthesizer using a DMT-ON protocol on a 1 µmol scale (1000 Å CPG column). Amine-modifier C6 dA (Glen Research 10-1089), dA+dC+dG+dT-CE Phosphoramindite (Glen Research 10-1000, 10-1010, 10-1020, 10-1030), Chemical Phosphorylation Reagent II (10-1901) were incorporated as specified by the manufacturer. Following synthesis, the oligonucleotide was cleaved from the resin by incubation at 25°C in 400 µL of a 1:1 mixture of ammonium hydroxide and 40% methylamine for 25 minutes. The cleaved resin was filtered away by filtration, followed by incubation at 60°C for 30 minutes to remove the protecting groups on phosphoramidites. The oligonucleotide was concentrated under reduced pressure using a speedvac. The residue was then taken up into 100 µL of H_2O, and purified using reverse-phase HPLC purification using a [10% acetonitrile in 0.1 M TEAA, pH7] to [80% acetonitrile in 0.1 M TEAA, pH7] solvent gradient with a column temperature of 45°C. The purified oligonucleotide was then incubated at room temperature in 400 µL of 40% aqueous acetic acid for 1 h to cleave the DMT group, and then frozen and lyophilized. The oligonucleotide was incubated in 500 µL 30% ammonium hydroxide at room temperature for 15 minutes to cleave the CPRII linker. Following deprotection, the oligonucleotide was concentrated under reduced pressure using a speedvac. The dried product was dissolved into 100 µL H_2O and subjected to reverse-phase HPLC purification using a [10% acetonitrile in 0.1M TEAA, pH 7] to [80% acetonitrile in 0.1M TEAA, pH 7] solvent gradient with a column temperature of 45°C. The purified oligonucleotide was dissolved in water.

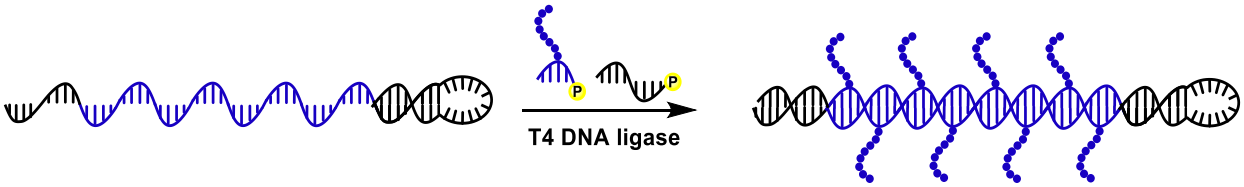
Functionalization of amino modified pentanucleotides with octa-peptide through DST linker

A mixture of 5 µL amino modified pentanucleotides (2000 µM in H_2O), 5 µL Na_2CO_3 buffer (pH 9.7, 500mM), 5 µL bis(2,5-dioxopyrrolidin-1-yl) 2,3-dihydroxysuccinate (DST, 100 mM in DMSO) and 35 µL DMSO was incubated at room temperature for 90 seconds with vortex. Followed by addition of 5 µL of N terminal free octa-peptide (50mM in DMSO). The mixture was incubated at room temperature for 2 hours under vortex. The reaction was then watered up to 100 µL and subjected to reverse-phase HPLC using a [10% acetonitrile in 0.1M TEAA, pH 7] to [80% acetonitrile in 0.1M TEAA, pH 7] solvent gradient with a column temperature of 45°C. The purified pentanucleotides were dissolved in water and characterized with mass spectrometry.

Pentanucleotide	**Sequence**	**Calculated mass range**	**Observed mass range**
PMeltT-pep	<P>ACTGT	2642.25	2643.0
PA16-1-pep	<P>ASWSW	2619.35-2717.39	2620.0-2717.0
PA16-2-pep	<P>AWSWS	2619.35-2717.39	2620.3-2716.4
PA32-pep	<P>AWWNS	2619.35-2717.39	2618.3-2719.2
PA64-pep	<P>AWNNS	2604.35-2733.38	2602.3-2734.3
PA16-2-DST	<P>AWSWS	1763.37-1861.41	1764.6-1866.6
PA16-2-CHO	<P>AWSWS	1687.36-1785.39	n.d.

Table S3.1: Mass spectrometric characterization of modified pentanucleotides

T4-DNA Ligase-mediate polymerization



Scheme S3.1: T4 DNA ligase mediated proliferation with peptide modification

In a PCR tube was added DNA template (15 pmol in 1.5 µL of water), primer (22.5 pmol in 2.25 µL of water) 10µL of ligation buffer (132 mM Tris-HCl, 20 mM $MgCl_2$, 2 mM dithiothreitol, 20% polyethylene glycol (PEG 6000), pH7.6), 1.75 µL of water and ATP (0.5nmol in 0.5 µL of water). The reaction mixture was heated to 90 °C for 2 minutes and then cooled down to 25 °C at the rate of 0.1 °C/s. In this PCR tube was then added functionalized pentanucleotides (480 pmol in 1 µL of water; 4 equivalents/codon), BSA (2 µg in 1 µL of water), 800 U of T4 DNA ligase (New England Biolabs, M0202L). The polymerization was performed at 25 °C for 24 h and then purified with MinElute® Reaction Cleanup Kit. The crude polymerized material was separated for analysis using denaturing PAGE (15% TBE, 150 V, 55 °C) and then stained with ethidium bromide and imaged by UV illumination.

Removal of DST-linked peptide on polymerized product

In a PCR tube was added purified polymerization product (10 pmol in 10 µL of water), and 90µL $NaIO_4$(55mM in water) solution. The reaction mixture was incubated under 25^{o}C for 2 hours and then desalted by gel filtration using CENTRI • SEP Spin Columns (Princeton Separations) equilibrated with water. $NaBH_4$ reduction was performed following that. In a PCR tube was added purified cleaved product (8 pmol in 10 µL of water), 30 µL of $NaBH_4$(1.33M in methanol) and 60 µL of water. Reaction mixture was

incubated ate 25ºC for 30 minutes followed by quenching with 20 µL sodium acetate (0.3M, pH 5 in water) for 15 minutes. Reduced product was desalted by gel filtration using CENTRI • SEP Spin Columns (Princeton Separations) equilibrated with water. The crude product was separated for analysis using denaturing PAGE (15% TBE, 150 V, 55 °C) and then stained with ethidium bromide and imaged by UV illumination.

Restriction enzyme digestion

In a PCR tube was added reduction product (5 pmol in 10 µL water), 5 µL 10X Cut smart buffer (New England Biolabs, B7204S), 10U of Hpy188I(for ASWSW library) (New England Biolabs, R0617S) or 10U of HpyCH4III(for AWSWS, AWWNS and AWNNS library)(New England Biolabs, R0618S)and water up to 50 µL . The digestion was carried out under 37ºC for 16 hours followed by MinElute® Reaction Cleanup Kit. The crude product was separated for analysis using denaturing PAGE (15% TBE, 150 V, 55 °C) and then stained by ethidium bromide and imaged by UV illumination.

Adapter duplex synthesis

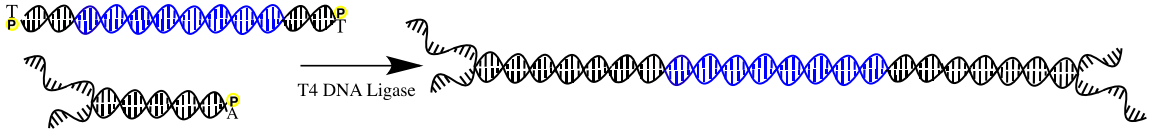
Scheme S3.2: Generation of adapter duplex for ligation

In a PCR tube was added 15 µL of 100 µM adapterA and 15 µL of 100 µM adapterB, then the tube was heated to 95 °C for 5 minutes and cooled to room temperature over one hour. Then in this PCR tube was added 4 µL NEBuffer2 10× (New England Biolabs, M0212L), 25 U Klenow Fragment (3'→5' exo$^-$, New England Biolabs, M0212L), 1 µL dNTP Mix (Thermo Scientific, 10 mM each). The extension was performed at 37 °C for

1 hour. The adapter duplex was purified with QIAquick Nucleotide Removal Kit, and then diluted in 30 µL water.

In a PCR tube was added 30 µL purified adapter duplex, 5 µl NEBuffer2 10× (New England Biolabs, M0212L), 25 U Klenow Fragment (3'→5' exo$^-$, New England Biolabs, M0212L), 5 µL dATP (Thermo Scientific, 10 mM), 5 µL H_2O. This PCR tube was incubated at 37 °C for 1 hour for A-tailing. Then product was purified with QIAquick Nucleotide Removal Kit, and then diluted in 30 µL water.

Adapter ligation



Scheme S3.3: Ligation of the Adapter duplex to dsDNA for sequencing

In a PCR tube was added 10pmol polymerization products, 200pmol A-tailing adapter duplex, 10µL NEBNext® Quick Ligation Reaction Buffer 5X, 2.5µL BSA (2mg/mL in H_2O), 1000U T4 DNA ligase (New England Biolabs, M0202L), the total volume of reaction was adjusted to 50µL with H_2O. Then the ligation was performed at 16 °C for 16 hours. The ligated products were then gel purified.

PCR protocols for sequencing

Each purified adapter ligation product was amplified with a different primer from iTrus_D701 to iTrus_D704.

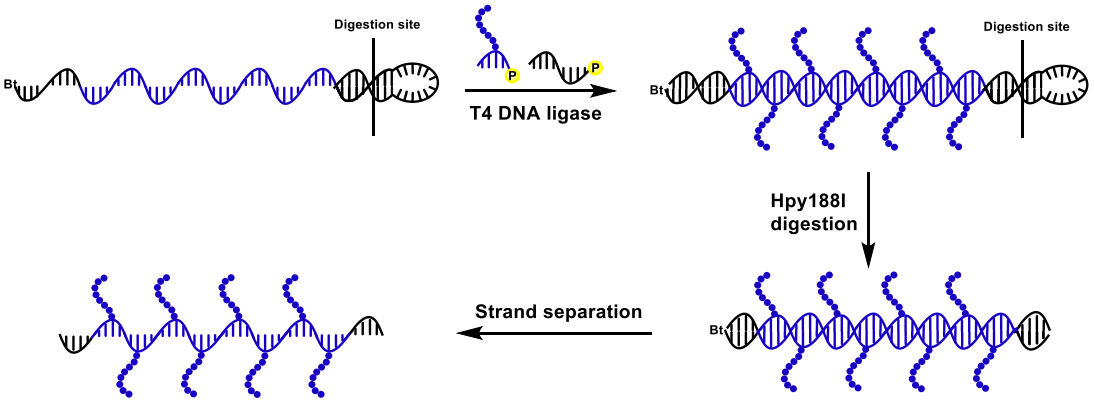
In a PCR tube was added 50 attomole purified adapter ligation product in 10 µL H_2O, 1.25µL 10 µM Primer B, 1.25 µL 10 µM corresponding iTrus_D7XX primer, and 12.5 µL Q5® High-Fidelity 2× Master Mix (New England Biolabs). The tube was then transferred to a preheated themocylcler (98°C). The PCR cycle was started with 10 s of

98°C denature step, followed by 30 s of primer annealing step (annealing temperature was 55°C for the first two cycles, and 71 °C for the rest of the cycles), and 30 s of 72°C extension step. The PCR products were then gel purified.

High-Throughput DNA sequencing Protocol

The concentrations of gel purified samples were determined with Kapa library quantification kit for Illumina libraries (KK4845) on Roche LightCycler 480. Paired-end Illumina sequencing was performed on an Illumina MiSeq system using the kit v2 with 300 cycles (150bp PE sequencing) at the Georgia Genomics Facility, University of Georgia, Athens, GA, USA

Preparation of octa-peptide modified MeltT



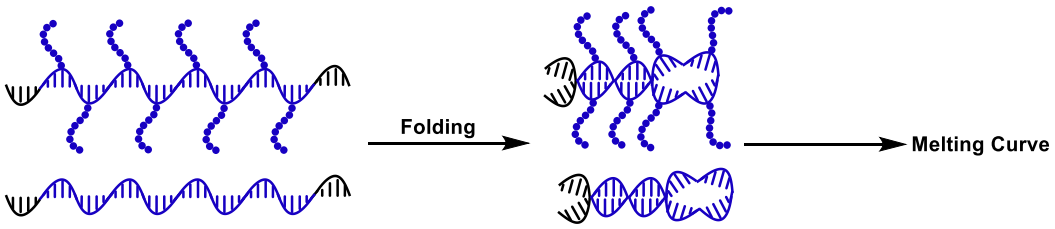
Scheme S3.4: Preparation of octa-peptide modified MeltT for thermal stable testing

Polymerization with 150 pmol scale was carried out by following the above protocol using MeltT template, MeltT primer and Melt-peptide pentanucleotide instead. Followed by Hpy188I restriction enzyme digestion to generate dsDNA (~40pmol collected).

In a PCR strip was added 20 µL Dynabeads® MyOne™ Streptavidin C1(Thermo Fisher Scientific, 65001). Wash the beads three times with 1X binding buffer(10mM Tris, 1M NaCl, 1mM EDTA and 0.05%v/v Tween®-20, pH 7.6). Resuspand beads in 40µL 2X

binding buffer followed by addition of 40μL dsDNA solution. Incubate the mixture under room temperature for 15 minutes. Removed the solution with magnetic rack and washed the beads twice with 1X binding buffer. Non-biotinylated strand was eluted with 40μL NaOH(150mM) incubation under 37°C for 1 minute. Took the supernatant out and neutralized with 4μL HCl(1.5M). The ssDNA was desalted with CENTRI•SEP Spin Column (Princeton Separations) and the concentration was determined by Nanodrop 2000 spectrophotometer (Thermo Fisher Scientific).

Melting curve of natural and octa-peptide modified ssDNA



Scheme S3.5: Thermal melting against native and modified ssDNA

10pmol of natural/peptide modified ssDNA was dissolved in 50μL of 1X sodium citrate buffer (0.15M NaCl, 0.015 sodium citrate pH 7) to prepare the folding buffer. Heat the folding solution at 90°C for 5 minutes and cool down to room temperature overnight. After fully folding, 0.5μL of 100X sybr green (Thermo Fisher Scientific, S7563) was added together with 49.5μL of folding solution to prepare the testing solution. Melting curve was achieved on MiniOpticon Real Time PCR System (Bio-rad) by incubating the testing solution under the melting protocol which heats from 25°C to 95°C with an increment of 0.5C and a plate reading of 5s.

3.6 Supporting results

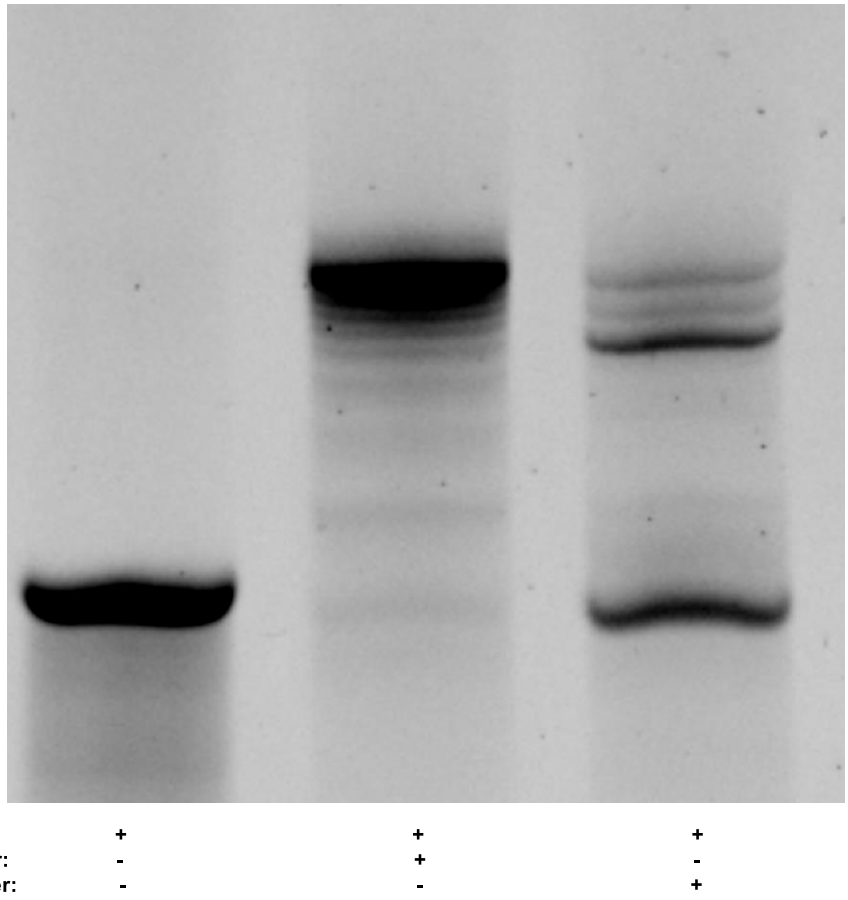


Figure S3.2: Inhibition of polymerization efficiency with aldehyde functional group. Polymerizations were performed using a 3'-hairpin, 5'-primer containing template comprising a reading frame of 8 randomized codons from the WSWST codon set. The peptide conjugated to the pentanucleotide library was GSASIFLY, which was attached via a DST linker. The CHO (aldehyde) pentanucleotide library was prepared by oxidative cleavage of the DST linker.

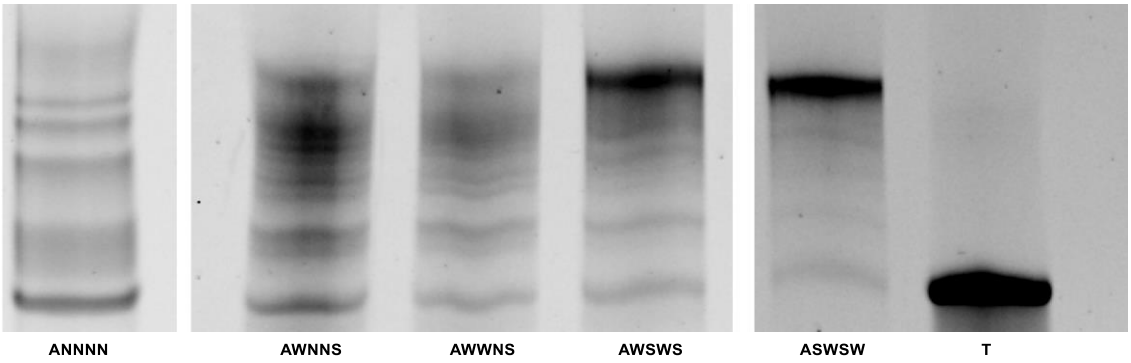


Figure S3.3: LOOPER efficiency with various GSASIFLY-conjugated pentanucleotide libraries along corresponding library of template comprising a reading frame of eight codons. Note that polymerization efficiency falls sharply after codon set size is expanded beyond 16-members. N = A, C, G, T mixture; S = C, G mixture; W = A, T mixture.

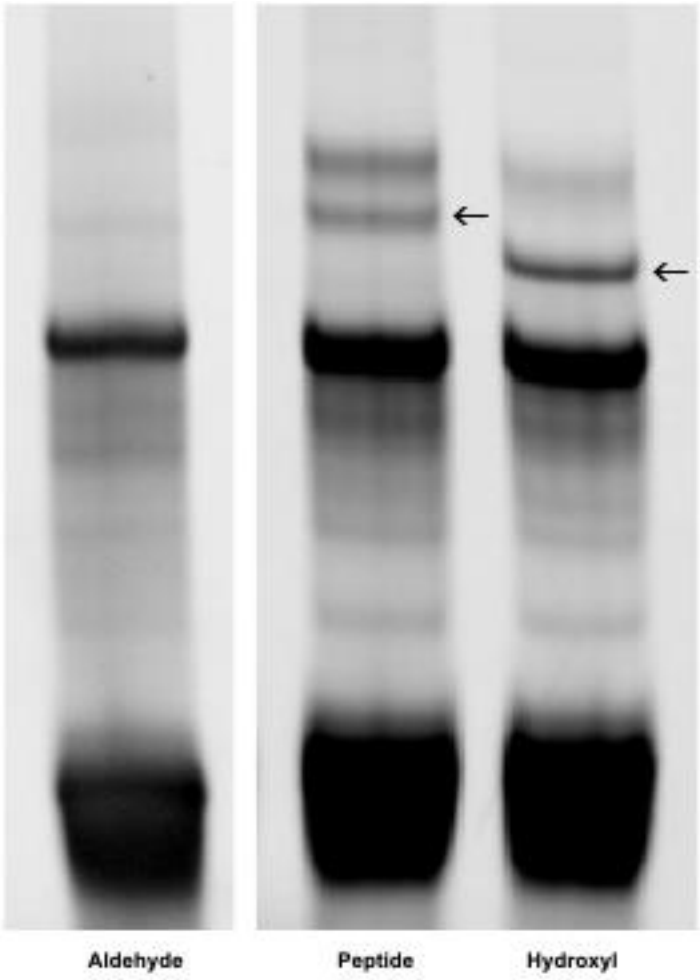


Figure S3.4: Ligation efficiency of the adapter duplex installation. 14N bp duplex sequencing adapters were installed on various duplex products of LOOPER. Arrows indicated desired double ligation band. Ligation of duplex product containing aldehyde motifs failed to yield sufficient material for illumina sequencing.

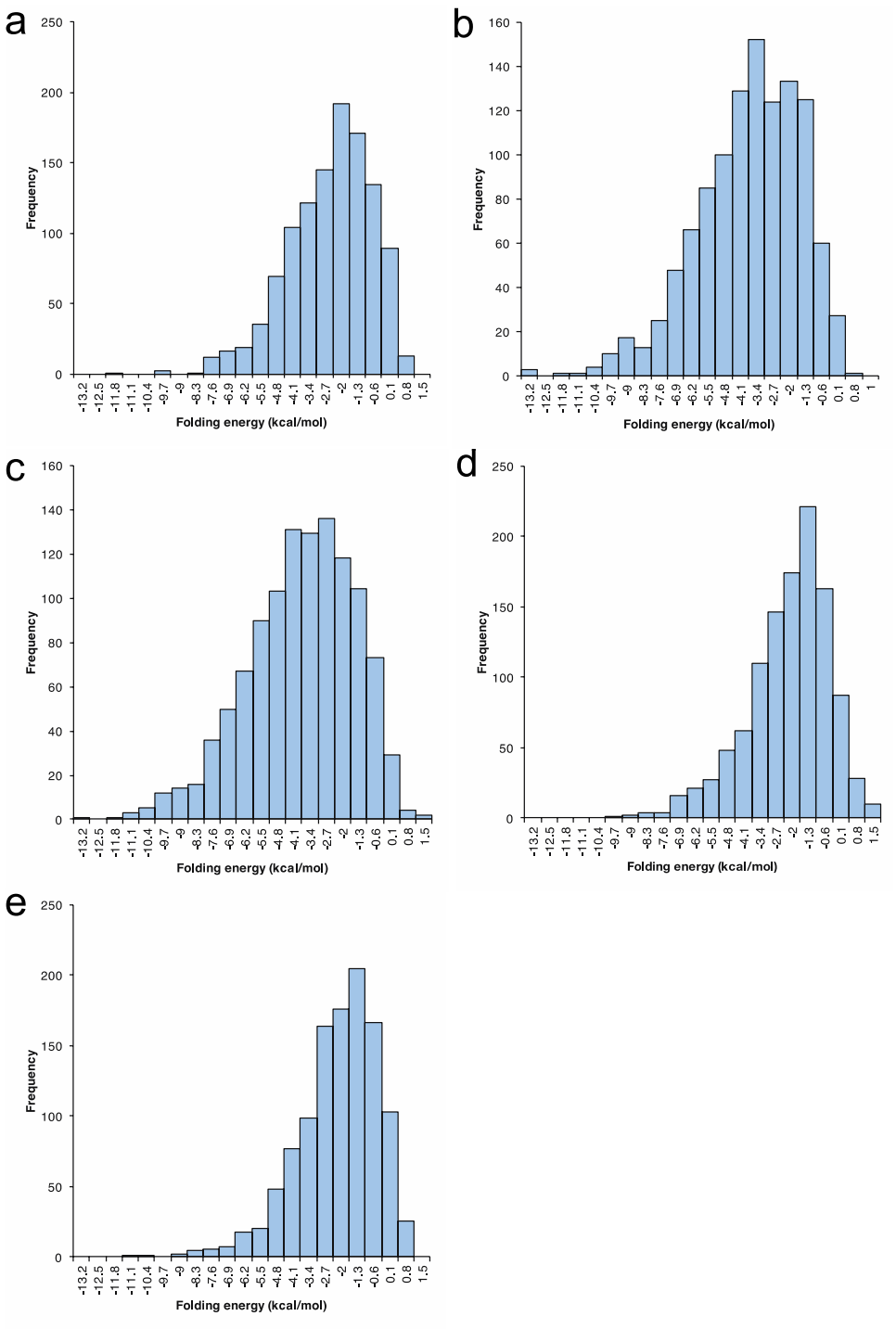


Figure S3.5: Modelling of folding thermodynamics for various DNA libraries. 1,125 library members for

a) $(NNNNN)_8$; b) $(ASWSW)_8$; c) $(AWSWS)_8$; d) $(AWWNS)_8$; e) $(AWNNS)_8$ were randomly computer generated and the Gibbs energy of folding calculated by Quickfold batch function using the DINAMelt web server.

Table S3.2 Fidelity of individual pentanucleotides during LOOPER

Entry	Pentanucleotide[a]	Codon	Enrichment[b]	Fidelity[c]
1	5'P-A*TGTG	CACAT	1.15	99.6%
2	5'P-A*AGTG	CACTT	0.94	99.5%
3	5'P-A*TCTG	CAGAT	1.30	99.5%
4	5'P-A*ACTG	CAGTT	0.75	99.4%
5	5'P-A*TGAG	CTCAT	1.06	99.3%
6	5'P-A*AGAG	CTCTT	0.77	99.2%
7	5'P-A*TCAG	CTGAT	1.18	99.1%
8	5'P-A*ACAG	CTGTT	0.99	99.1%
9	5'P-A*TGTC	GACAT	1.09	99.0%
10	5'P-A*AGTC	GACTT	1.04	99.0%
11	5'P-A*TCTC	GAGAT	1.13	98.9%
12	5'P-A*ACTC	GAGTT	0.99	98.9%
13	5'P-A*TGAC	GTCAT	1.05	98.5%
14	5'P-A*AGAC	GTCTT	0.76	98.4%
15	5'P-A*TCAC	GTGAT	1.07	97.7%
16	5'P-A*ACAC	GTGTT	1.04	96.6%

Entry	Pentanucleotide[a]	Codon	Enrichment[b]	Fidelity[c]
1	5’P-A*TTTG	CAAAT	0.94	98.8%
2	5’P-A*ATTG	CAATT	0.73	98.4%
3	5’P-A*TATG	CATAT	0.97	97.5%
4	5’P-A*AATG	CATTT	0.57	97.1%
5	5’P-A*TTGG	CCAAT	1.29	99.3%
6	5’P-A*ATGG	CCATT	0.92	98.0%
7	5’P-A*TAGG	CCTAT	1.08	98.7%
8	5’P-A*AAGG	CCTTT	0.58	97.4%
9	5’P-A*TTCG	CGAAT	1.74	99.1%
10	5’P-A*ATCG	CGATT	0.99	98.5%
11	5’P-A*TACG	CGTAT	1.34	99.0%
12	5’P-A*AACG	CGTTT	0.76	98.0%
13	5’P-A*TTAG	CTAAT	1.45	98.6%
14	5’P-A*ATAG	CTATT	0.84	96.8%
15	5’P-A*TAAG	CTTAT	1.08	98.0%
16	5’P-A*AAAG	CTTTT	0.52	97.6%
17	5’P-A*TTTC	GAAAT	1.36	98.9%
18	5’P-A*ATTC	GAATT	0.93	98.3%
19	5’P-A*TATC	GATAT	1.09	98.2%

20	5’P-A*AATC	GATTT	0.61	93.8%
21	5’P-A*TTGC	GCAAT	1.36	99.3%
22	5’P-A*ATGC	GCATT	0.95	98.4%
23	5’P-A*TAGC	GCTAT	1.16	98.7%
24	5’P-A*AAGC	GCTTT	0.65	97.0%
25	5’P-A*TTCC	GGAAT	2.04	98.9%
26	5’P-A*ATCC	GGATT	1.46	98.3%
27	5’P-A*TACC	GGTAT	1.53	98.5%
28	5’P-A*AACC	GGTTT	1.10	97.7%
29	5’P-A*TTAC	GTAAT	1.41	98.3%
30	5’P-A*ATAC	GTATT	0.94	97.1%
31	5’P-A*TAAC	GTTAT	1.06	97.6%
32	5’P-A*AAAC	GTTTT	0.62	97.3%

Entry	Pentanucleotide[a]	Codon	Enrichment[b]	Fidelity[c]
1	5’P-A*TTTG	CAAAT	1.08	97.0%
2	5’P-A*ATTG	CAATT	0.76	95.6%
3	5’P-A*TGTG	CACAT	0.90	92.3%
4	5’P-A*AGTG	CACTT	0.50	88.3%
5	5’P-A*TCTG	CAGAT	1.53	97.0%
6	5’P-A*ACTG	CAGTT	0.84	93.4%

7	5’P-A*TATG	CATAT	0.96	94.1%
8	5’P-A*AATG	CATTT	0.57	94.2%
9	5’P-A*TTGG	CCAAT	1.39	97.6%
10	5’P-A*ATGG	CCATT	0.89	96.0%
11	5’P-A*TGGG	CCCAT	0.93	94.6%
12	5’P-A*AGGG	CCCTT	0.46	88.9%
13	5’P-A*TCGG	CCGAT	2.01	97.0%
14	5’P-A*ACGG	CCGTT	0.84	94.5%
15	5’P-A*TAGG	CCTAT	0.97	94.8%
16	5’P-A*AAGG	CCTTT	0.49	95.4%
17	5’P-A*TTCG	CGAAT	1.57	97.0%
18	5’P-A*ATCG	CGATT	0.98	96.6%
19	5’P-A*TGCG	CGCAT	0.81	96.3%
20	5’P-A*AGCG	CGCTT	0.58	87.9%
21	5’P-A*TCCG	CGGAT	1.96	95.6%
22	5’P-A*ACCG	CGGTT	1.21	93.8%
23	5’P-A*TACG	CGTAT	1.31	95.2%
24	5’P-A*AACG	CGTTT	0.74	95.9%
25	5’P-A*TTAG	CTAAT	1.43	96.2%
26	5’P-A*ATAG	CTATT	0.72	94.7%
27	5’P-A*TGAG	CTCAT	0.90	97.0%
28	5’P-A*AGAG	CTCTT	0.41	94.4%
29	5’P-A*TCAG	CTGAT	1.65	97.8%

30	5’P-A*ACAG	CTGTT	1.05	95.2%
31	5’P-A*TAAG	CTTAT	1.00	95.6%
32	5’P-A*AAAG	CTTTT	0.48	93.3%
33	5’P-A*TTTC	GAAAT	1.40	97.7%
34	5’P-A*ATTC	GAATT	0.96	94.6%
35	5’P-A*TGTC	GACAT	0.98	97.0%
36	5’P-A*AGTC	GACTT	0.58	94.9%
37	5’P-A*TCTC	GAGAT	1.65	96.0%
38	5’P-A*ACTC	GAGTT	1.29	94.7%
39	5’P-A*TATC	GATAT	1.01	94.2%
40	5’P-A*AATC	GATTT	0.65	87.5%
41	5’P-A*TTGC	GCAAT	1.19	96.8%
42	5’P-A*ATGC	GCATT	0.88	97.2%
43	5’P-A*TGGC	GCCAT	0.93	96.9%
44	5’P-A*AGGC	GCCTT	0.57	91.6%
45	5’P-A*TCGC	GCGAT	1.57	96.3%
46	5’P-A*ACGC	GCGTT	0.95	92.9%
47	5’P-A*TAGC	GCTAT	1.05	94.3%
48	5’P-A*AAGC	GCTTT	0.57	94.5%
49	5’P-A*TTCC	GGAAT	1.77	97.2%
50	5’P-A*ATCC	GGATT	1.26	96.4%
51	5’P-A*TGCC	GGCAT	1.36	97.0%
52	5’P-A*AGCC	GGCTT	0.66	92.1%

53	5’P-A*TCCC	GGGAT	2.27	95.1%
54	5’P-A*ACCC	GGGTT	1.60	94.2%
55	5’P-A*TACC	GGTAT	1.44	94.7%
56	5’P-A*AACC	GGTTT	0.96	94.8%
57	5’P-A*TTAC	GTAAT	1.28	95.9%
58	5’P-A*ATAC	GTATT	0.76	94.2%
59	5’P-A*TGAC	GTCAT	0.98	96.6%
60	5’P-A*AGAC	GTCTT	0.44	91.5%
61	5’P-A*TCAC	GTGAT	1.60	96.4%
62	5’P-A*ACAC	GTGTT	1.24	95.8%
63	5’P-A*TAAC	GTTAT	0.97	94.6%
64	5’P-A*AAAC	GTTTT	0.57	94.0%

[a] pentanucleotides were modified with GSASIFLY at 5’dA, denoted by an asterisk. [b] Enrichment was calculated as (freq. polymer/freq. template). [c]Fidelity was calculated for pentanucleotide incorporation; the fidelity is considerably higher if evaluated at the single nucleotide level.

3.7 References:

1. Jones, K. A.; Kadonaga, J. T.; Rosenfeld, P. J.; Kelly, T. J.; Tjian, R., A cellular DNA-binding protein that activates eukaryotic transcription and DNA replication. *Cell* **1987,** *48* (1), 79-89.

2. Bode, A. M.; Dong, Z., Post-translational modification of p53 in tumorigenesis. *Nat Rev Cancer* **2004,** *4* (10), 793-805.

3. Burgering, B. M.; Coffer, P. J., Protein kinase B (c-Akt) in phosphatidylinositol-3-OH kinase signal transduction. *Nature* **1995,** *376* (6541), 599-602.

4. Jones, S.; Thornton, J. M., Principles of protein-protein interactions. *Proc Natl Acad Sci U S A* **1996,** *93* (1), 13-20.

5. Ryan, D. P.; Matthews, J. M., Protein-protein interactions in human disease. *Curr Opin Struct Biol* **2005,** *15* (4), 441-6.

6. Robinson, J. A.; Demarco, S.; Gombert, F.; Moehle, K.; Obrecht, D., The design, structures and therapeutic potential of protein epitope mimetics. *Drug Discov Today* **2008,** *13* (21-22), 944-51.

7. Arkin, M. R.; Tang, Y.; Wells, J. A., Small-molecule inhibitors of protein-protein interactions: progressing toward the reality. *Chem Biol* **2014,** *21* (9), 1102-14.

8. Azzarito, V.; Long, K.; Murphy, N. S.; Wilson, A. J., Inhibition of alpha-helix-mediated protein-protein interactions using designed molecules. *Nat Chem* **2013,** *5* (3), 161-73.

9. (a) Wilson, D. S.; Szostak, J. W., In vitro selection of functional nucleic acids. *Annu Rev Biochem* **1999,** *68*, 611-47; (b) Packer, M. S.; Liu, D. R., Methods for the directed evolution of proteins. *Nat Rev Genet* **2015,** *16* (7), 379-94.

10. (a) Lou, C.; Martos-Maldonado, M. C.; Madsen, C. S.; Thomsen, R. P.; Midtgaard, S. R.; Christensen, N. J.; Kjems, J.; Thulstrup, P. W.; Wengel, J.; Jensen, K. J., Peptide-oligonucleotide conjugates as nanoscale building blocks for assembly of an artificial three-helix protein mimic. *Nat Commun* **2016,** *7*, 12294; (b) Ghosh, P. S.;

Hamilton, A. D., Noncovalent template-assisted mimicry of multiloop protein surfaces: assembling discontinuous and functional domains. *J Am Chem Soc* **2012,** *134* (32), 13208-11; (c) Diezmann, F.; von Kleist, L.; Haucke, V.; Seitz, O., Probing heterobivalent binding to the endocytic AP-2 adaptor complex by DNA-based spatial screening. *Org Biomol Chem* **2015,** *13* (29), 8008-15; (d) Dix, A. V.; Conroy, J. L.; George Rosenker, K. M.; Sibley, D. R.; Appella, D. H., PNA-Based Multivalent Scaffolds Activate the Dopamine D2 Receptor. *ACS Med Chem Lett* **2015,** *6* (4), 425-9; (e) Dix, A. V.; Moss, S. M.; Phan, K.; Hoppe, T.; Paoletta, S.; Kozma, E.; Gao, Z. G.; Durell, S. R.; Jacobson, K. A.; Appella, D. H., Programmable nanoscaffolds that control ligand display to a G-protein-coupled receptor in membranes to allow dissection of Diezmann, F.; Seitz, O., DNA as a molecular ruler: interrogation of a tandem SH2 domain with self-assembled, bivalent DNA-peptide complexes. *Angew Chem Int Ed Engl* **2011,** *50* (18), 4146-50; (g) Englund, E. A.; Wang, D.; Fujigaki, H.; Sakai, H.; Micklitsch, C. M.; Ghirlando, R.; Martin-Manso, G.; Pendrak, M. L.; Roberts, D. D.; Durell, S. R.; Appella, D. H., Programmable multivalent display of receptor ligands using peptide nucleic acid nanoscaffolds. *Nat Commun* **2012,** *3*, 614; (h) Finke, A.; Busskamp, H.; Manea, M.; Marx, A., Designer Extracellular Matrix Based on DNA-Peptide Networks Generated by Polymerase Chain Reaction. *Angew Chem Int Ed Engl* **2016**; (i) Flory, J. D.; Shinde, S.; Lin, S.; Liu, Y.; Yan, H.; Ghirlanda, G.; Fromme, P., PNA-peptide assembly in a 3D DNA nanocage at room temperature. *J Am Chem Soc* **2013,** *135* (18), 6985-93; (j) Gorska, K.; Huang, K. T.; Chaloin, O.; Winssinger, N., DNA-Templated Homo- and Heterodimerization of Peptide Nucleic Acid Encoded Oligosaccharides that Mimick the Carbohydrate Epitope of HIV. *Angew Chem Int Edit* **2009,** *48* (41), 7695-7700; (k)

Humenik, M.; Scheibel, T., Nanomaterial Building Blocks Based on Spider Silk-Oligonucleotide Conjugates. *Acs Nano* **2014,** *8* (2), 1342-1349; (l) Janssen, B. M. G.; Lempens, E. H. M.; Olijve, L. L. C.; Voets, I. K.; van Dongen, J. L. J.; de Greef, T. F. A.; Merkx, M., Reversible blocking of antibodies using bivalent peptide-DNA conjugates allows protease-activatable targeting. *Chem Sci* **2013,** *4* (4), 1442-1450; (m) Kazane, S. A.; Axup, J. Y.; Kim, C. H.; Ciobanu, M.; Wold, E. D.; Barluenga, S.; Hutchins, B. A.; Schultz, P. G.; Winssinger, N.; Smider, V. V., Self-assembled antibody multimers through peptide nucleic acid conjugation. *J Am Chem Soc* **2013,** *135* (1), 340-6; (n) Kovacic, S.; Samii, L.; Lamour, G.; Li, H.; Linke, H.; Bromley, E. H.; Woolfson, D. N.; Curmi, P. M.; Forde, N. R., Construction and characterization of kilobasepair densely labeled peptide-DNA. *Biomacromolecules* **2014,** *15* (11), 4065-72; (o) Liang, S. I.; McFarland, J. M.; Rabuka, D.; Gartner, Z. J., A modular approach for assembling aldehyde-tagged proteins on DNA scaffolds. *J Am Chem Soc* **2014,** *136* (31), 10850-3; (p) Welter, M.; Verga, D.; Marx, A., Sequence-Specific Incorporation of Enzyme-Nucleotide Chimera by DNA Polymerases. *Angew Chem Int Ed Engl* **2016**; (q) Williams, B. A.; Diehnelt, C. W.; Belcher, P.; Greving, M.; Woodbury, N. W.; Johnston, S. A.; Chaput, J. C., Creating protein affinity reagents by combining peptide ligands on synthetic DNA scaffolds. *J Am Chem Soc* **2009,** *131* (47), 17233-41.

11. Guo, C.; Watkins, C. P.; Hili, R., Sequence-Defined Scaffolding of Peptides on Nucleic Acid Polymers. *J Am Chem Soc* **2015,** *137* (34), 11191-6.

12. Lei, Y.; Kong, D.; Hili, R., A High-Fidelity Codon Set for the T4 DNA Ligase-Catalyzed Polymerization of Modified Oligonucleotides. *ACS Comb Sci* **2015,** *17* (12), 716-21.

13. Markham, N. R.; Zuker, M., DINAMelt web server for nucleic acid melting prediction. *Nucleic Acids Res* **2005,** *33* (Web Server issue), W577-81.

14. Kong, D.; Lei, Y.; Yeung, W.; Hili, R., Enzymatic Synthesis of Sequence-Defined Synthetic Nucleic Acid Polymers with Diverse Functional Groups. *Angew Chem Int Ed Engl* **2016,** *55* (42), 13164-13168.

Chapter 4

Influence of linker length on ligase-catalyzed oligonucleotide polymerization[1]

1. Guo, C.; Mahdavi, A, Y.; Hili, R.; To be submitted to *ChembioChem*.

4.1 Abstract

Ligase-catalyzed oligonucleotide polymerization (LOOPER) has been recently developed to enable the sequence-defined generation of DNA with up to 16 different modifications. This approach was used to evolve new classes of diversely-modified DNA aptamers for molecular recognition. The modifications in LOOPER are appended via a long hexamethylenediamine linker, which could negatively impact binding thermodynamics. Herein we explore the incorporation of modifications via shorter linkers using commercially available phosphoramidites and assess their efficiency and fidelity of incorporation. We observed that shorter linkers are less tolerated during LOOPER, with very short linkers providing high levels of error and sequence bias. An ethylenediamine linker was found to be optimal in terms of yield, efficiency, and bias; however, codon adjustment was necessary. This shorter linker anticodon set for LOOPER should prove valuable in exploring the impact of diverse chemical modifications on the molecular function of DNA.

4.2 Introduction

Nucleic acid polymers can be readily evolved by in vitro selection methods to engender molecular function, such as molecular recognition and catalysis.[1-2] Due to their chemical stability and reproducible folding, they have often been adopted in place of antibodies as molecular recognition reagents for diagnostics and have demonstrated promise as therapeutics.[3-4] The ability of nucleic acids to bind a wide variety of targets ranging from metal ions[5] to cells[6] is remarkable considering their limited chemical diversity. However, increasing the chemical diversity of nucleic acid polymers has been a long-sought

challenge in order to expand their functional properties.[7-8] The conventional strategy for sequence-defined incorporation of modifications throughout a nucleic acid polymer is to use polymerase-catalyzed primer extensions with modified (deoxy)nucleotide triphosphates. Using a standard four base-pair genetic code, this can permit the incorporation of up to four additional functional groups to be displayed. This approach has been used for homomultivalent and heteromultivalent display of functional groups to generate nucleic acid polymers with improved binding thermodynamics and nuclease stability versus canonical DNA. Recently, this approach was used to incorporate three different functional groups in DNA to expand its catalytic properties.[9]

Expanding beyond four modifications requires an alternative strategy that does not rely upon polymerase-mediated incorporation. To this end, Ligase-catalyzed OligOnucleotide PolymERization (LOOPER) permits the expansion of chemical diversity in DNA beyond four different modifications (**Figure 4.1**).[10-17] LOOPER involves the T4 DNA ligase-catalyzed copolymerization of a library of modified oligonucleotides along a library of DNA templates. Due to the high fidelity of T4 DNA ligases, the modified oligonucleotides (anticodons), which bear the chemical diversity, are incorporated across from their cognate codons with high efficiency and low error, resulting in a sequence-defined modified oligonucleotide product. The number of unique modifications that can be incorporated *via* LOOPER is dependent on two variables: (*i*) codon library size, defined as 4^n, where n is the codon length; and (*ii*) degeneracy, which is the redundancy within the codon set to define specific modifications. To date, the largest codon set used in LOOPER is defined as XXNNT, whereby XX encodes for one of 16 possible modifications on the cognate anticodon. Together, this codon sets comprises 256 codons,

and thus, each of the 16 modifications are encoded by 16 different codons. LOOPER has been successfully used in the in vitro selection of modified DNA aptamers. A modified thrombin aptamer was raised with a K_d of 1.6 nM, and represents the first thrombin DNA aptamer to break from the G-quadruplex archetype. Importantly, the modifications were found to be critical to molecular recognition of thrombin. More recent reports on the use of LOOPER to raise modified aptamers against other protein targets has further validated the approach.[18]

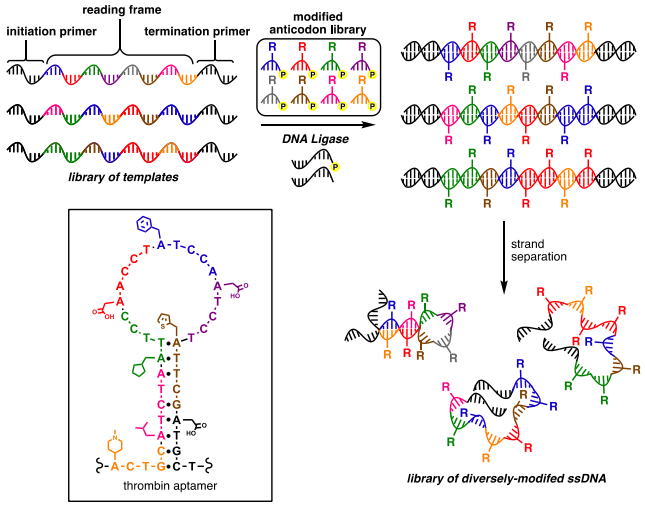


Figure 4.1: General process for generation of sequence-defined modified DNA for use during in vitro selection. Inset: example of modified aptamer raised against human α-thrombin.

The initial development of LOOPER and the first in vitro selection of a LOOPER-derived aptamer were reported using a hexamethylenediamine (HMDA) to bridge the 8-position of the adenine to variable functional groups installed by amide-bond coupling chemistry (**Figure 4.2**). HMDA is a long, flexible linker, which provides both advantages and disadvantages for ligand display. While a longer linker provides increased reach to engage in molecular recognition of the target, having excess linker length beyond optimal

geometry for binding decreases the effective molarity of the ligand resulting in decreased binding affinity.[19] Indeed, long linker lengths have been implicated in failed in vitro selections of modified aptamers.[20] Having ready access to shorter linker lengths that are well behaved in LOOPER will provide the opportunity to further improve the binding properties of modified aptamers derived from LOOPER. To evaluate the influence of linker length during LOOPER, we chose to compare the commercially available 8-modified HMDA dA phosphoramidite with two shorter linkers, ethylenediamine (EDA), and methylamine (MA) (**Figure 4.2**).

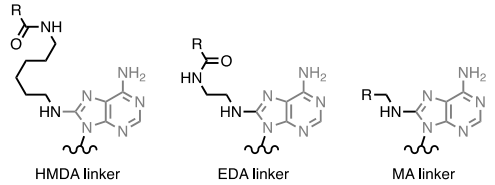


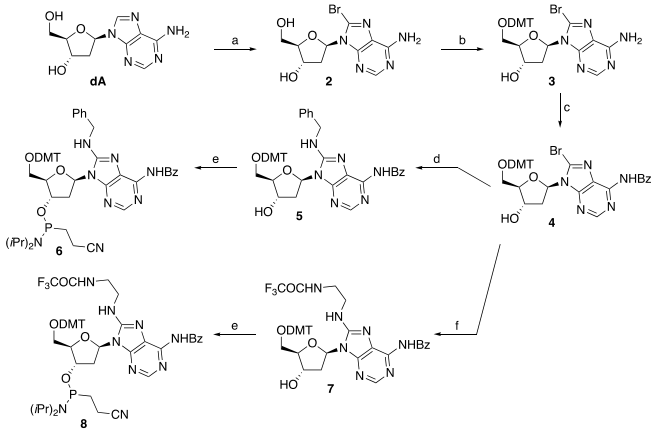
Figure 4.2: Linkers examined during LOOPER. HMDA = hexamethylenediamine; EDA = ethylenediamine; MA = methylamine.

4.3 Results and discussions

We began our study through the synthesis of phosphoramidite **6**, which included a benzylamine group at the 8-position of the adenine ring (**Scheme 4.1**). This would serve as a model for a minimally linked modification for LOOPER. The synthesis of **6** was adapted from a well-established synthetic route[21] that proceeds through 8-Br dA intermediate **4**, which can be subjected to nucleophilic displacement chemistry with various amine nucleophiles. This chemistry led to **6** in good yield and purity. Installation of **6** at the 5'position of a 256-membered A*NNNN anticodon library was achieved by automated DNA synthesis. The benzylamine-modified A*NNNN anticodon, **6a**, was

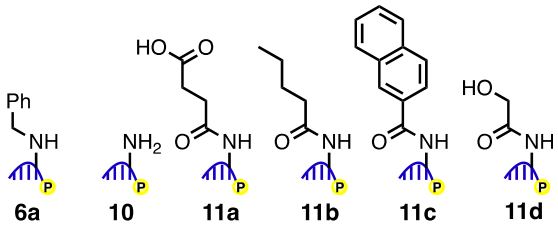
found to be well behaved in LOOPER giving 74% yield of full-length product when polymerized along a template comprising a reading frame with eight NNNNT codons. We next sought to evaluate the fidelity and codon bias for **6a** during LOOPER. To achieve this, we implemented a duplex DNA sequencing method previously reported by our lab,[14-15] which enables the high-throughput DNA sequencing of templates and corresponding LOOPER strands to determine the error rate at the single codon level. Using this technique, we found that **6a** polymerized with high fidelity (92.4%, **Table 4.1**, entry 1). When calculating the codon bias for the process, defined as the standard deviation of the frequency of a codon in the product divided by the frequency of the codon in template, we were surprised to see a high codon bias of 0.99. In previous studies,[14-15] we attributed poor codon bias to low yield, as some anticodons within specific codon sets were observed to polymerize at higher efficiencies; high codon bias resulting from poor yield usually became an issue below 60%. When examining the codon bias plot (**Figure 4.3a**), it became clear that several codons were inhibited by the modification during LOOPER. In the codon bias plot, large numbers of data points stray toward the y-axis, rather than along the x-y diagonal. Surprisingly, codons that suffered a negative bias were high in cytosine content; less surprising was the observation that those that experienced a positive bias were high in guanine content. For example, codons such as TCCCT and CCTCT suffered from a 10-fold decrease in codon frequency following LOOPER, while codons such as GGAGT and GGGGT were observed with 5 and 10-fold increases in codon frequency following LOOPER, respectively. The most striking result from the analysis was that large swaths of the codon set suffered from extremely high error, which is uncharacteristic of LOOPER. While the overall fidelity of the codon set

was high, 12 codons had error rates above 80%, 10 of which were over 90%; the codon TAAAT had an error rate of 100%. The exact mechanism by which a benzylamine-modified adenosine at the 5'-end of an anticodon results in such a drastic loss in fidelity during LOOPER is not currently understood. However, the data clearly suggest that the use of very small linkers to modifications could result in a significant loss of sequence diversity during the first round of in vitro selection.



Scheme 4.1: Synthesis of phosphoramidites **6** and **8**. a) Br_2, 1M acetate buffer, pH 4, r.t., 3 h, 69%. b) DMT-Cl, cat. DMAP, pyridine, r.t., 3 h, 86%. c) 1) TMS-Cl, pyridine, 0 °C - r.t., 1 h; 2) 5 eq. BzCl, 16 h; 3) NH_3 in MeOH, 5 °C, 1 h, 83 %. d) 5 eq. benzylamine, EtOH, 80 °C, 2 h, 79% e) 5.4 eq. DIPEA, 5.4 eq. 2-Cyanoethyl N,N-diisopropylchlorophosphoramidite, CH_2Cl_2, 0 °C, 3 h, (63% for **6**; 58% for **8**. f) 1) 5 eq. ethylenediamine, EtOH, 80 °C, 2 h; 2) ethyl trifluoroacetate, NEt_3, MeOH, 0 °C, 2 h, 81%.

Table 4.1: Fidelity and Bias analysis of LOOPER products[a]



Entry	Linker	Anticodon	Yield[b]	Fidelity[c]	Bias[d]
1	MA	6a	74%	92.4%	0.99
2	EDA	10	70%	93.4%	0.27
3	EDA	11a	77%	94.2%	0.50
4	EDA	11b	75%	94.1%	0.55
5	EDA	11c	67%	97.0%	0.49
6	EDA	11d	71%	92.3%	0.43

[a] LOOPER and duplex barcoding process performed on 15 pmol of DNA template library. 35 amol of product was subjected to duplex DNA sequencing. [b] Determined by densitometry of full-length product by polyacrylamide gel electrophoresis. [c] Fidelity was calculated for pentanucleotide incorporation. The fidelity at the single nucleotide level is higher due to the large majority of misincorporations having only one error per pentanucleotide [d] Codon bias was calculated as σ(freq. polymer/freq. template).

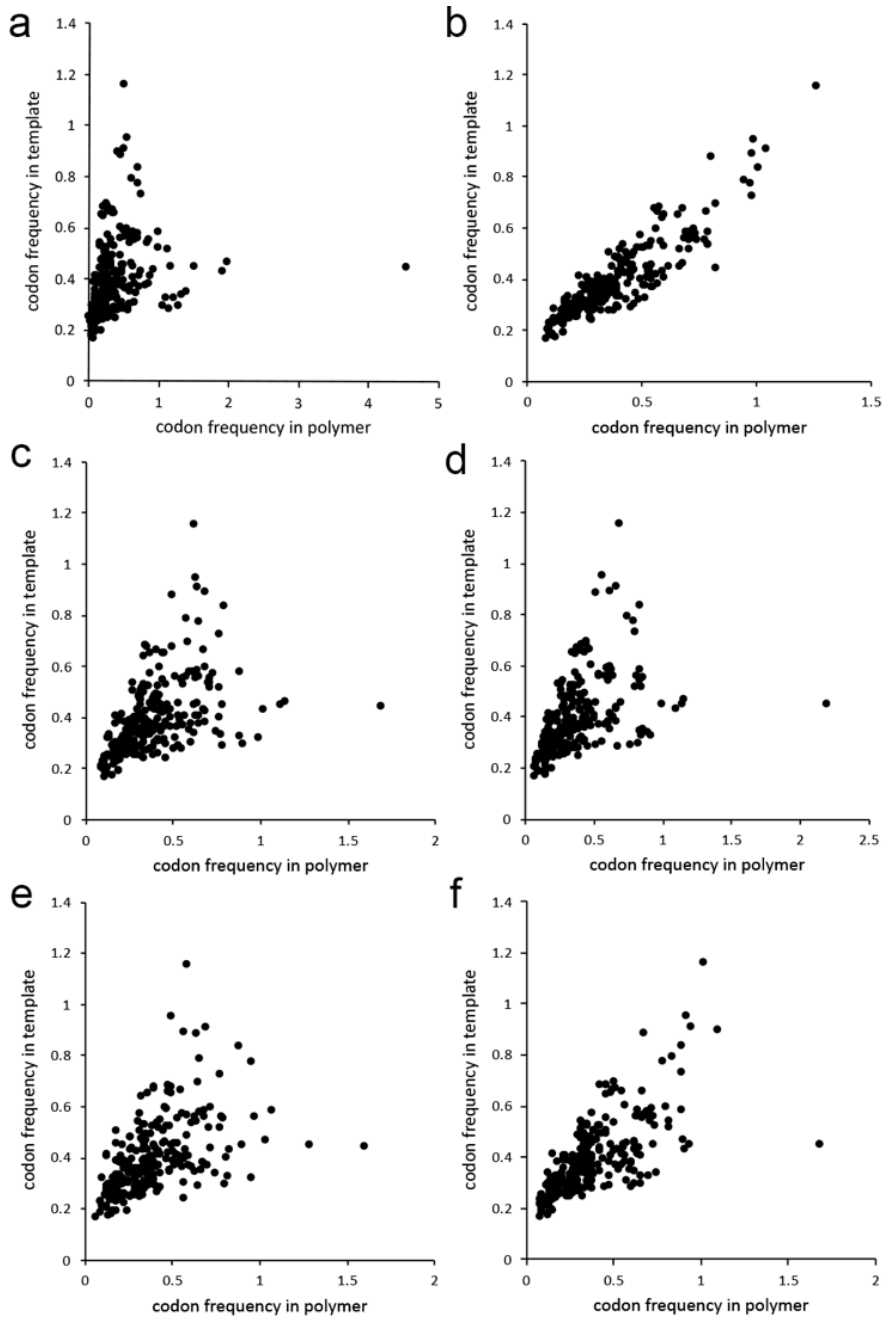
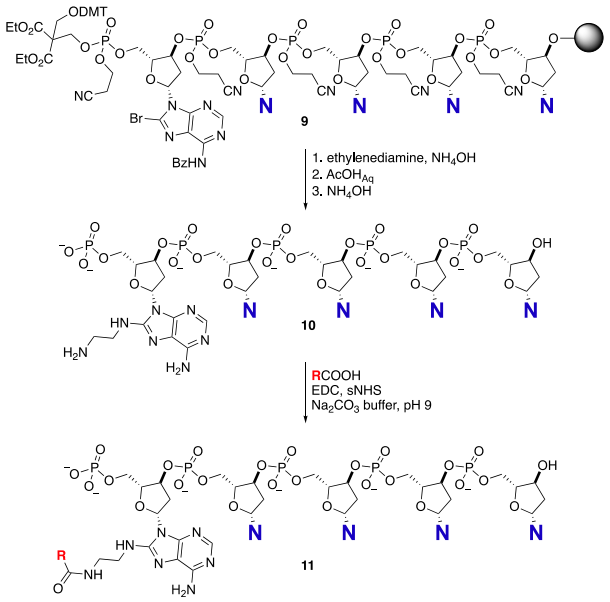


Figure 4.3: Bias analysis of NNNNT codon set during LOOPER using various different linker-modified ANNNN anticodons: (**a**) anticodon 6a; (**b**) anticodon 10; (**c**) anticodon 11a; (**d**) anticodon 11b; (**e**) anticodon 11c; (**f**) anticodon 11d.

We hypothesized that the large steric bulk close to the nucleobase was influencing hybridization and ligation kinetics; perhaps stabilizing the hybridization of mismatch incorporations for certain codon members. To address this, we evaluated a phosphoramidite that contained the slightly longer linker, ethylenediamine (EDA) (**Scheme 4.2**). The EDA-modified dA phosphoramidite would enable the incorporation of various functional groups via conventional EDC chemistry, which we previously used to modify the HMDA anticodons.[14] Thus, phosphoramidite **8** was prepared as a trifluoroacetamide protected amine via nucleophilic substitution on intermediate **4**. The incorporation of **8** into the 256-membered A*NNNN anticodon set was achieved by automated DNA synthesis and gave **10** in high yield, which was subsequently subjected to LOOPER along a DNA template comprising two primers flanking eight repeats of NNNNT. LOOPER proceeded efficiently at 70% yield of full-length product, and analysis of the product by duplex DNA sequencing revealed a 93.4% fidelity with a low codon bias of 0.27 (**Table 4.1**, entry 2). The codon bias plot showed a tight distribution along the horizontal, suggesting a well-behaved modified anticodon library. Indeed, the codons ranged in error between 1% and 15%, with the large majority under 10% error. We decided to pursue the EDA linker as a possible replacement for the HMDA linker by demonstrating its tolerance for a variety of chemical modifications. While the phosphoramidite synthesis of **8** was straightforward, we sought a convergent route that would minimize the requirement for phosphoramidite synthesis in order to facilitate its use in LOOPER. To this end, we used the commercially-available 8-Br dA phosphoramidite to generate A*NNNN libraries, **9**, by solid-phase automated DNA synthesis (**Scheme 4.2**). This would allow convergent synthesis of various modified

anticodons by direct nucleophilic substitution. Using excess EDA in the presence of ammonium hydroxide, we were able to synthesize the EDA-modified anticodon **10** in high yield and purity. Unfortunately, the method was not robust enough to accommodate mono-functionalized EDA due to the large excess required (10% volume) and the large steric bulk near the 8-Br dA in **9**. Thus, conversion of anticodon **10** into various modified anticodons **11a-d** was achieved by conventional DNA-compatible EDC-mediated amide coupling. Several derivatives of **11** were readily prepared to determine the scope of the EDA linker during LOOPER, including acids, amines, alcohols, aliphatics, and aromatics (**Table 4.1**).



Scheme 4.2: Synthesis of modified anticodon libraries containing an EDA linker.

Carboxylic acid modified anticodon, **11a**, was found to polymerize efficiently with high yield and fidelity with a modest codon bias (**Table 4.1**, entry 3). Analysis of the codon fidelity revealed that no codons resulted in error rates above 15%, suggesting that this modification is well behaved during LOOPER. Modification with an *n*-butyl chain resulted in similarly good yield and fidelities with modest codon bias; however,

sequencing analysis revealed that some codons had unacceptable error rates. Three anticodons polymerized with errors well above 15%, including ATTAT (32.4%), ATTTA (36.0%), and ATTTT (53.6%). While these all have 0% GC-content, this is not a determinant for poor fidelity; for example, AAATT polymerized with 5.6% error, while ATTTG polymerizes with only 1.4% error. Notwithstanding, should *n*-butyl groups be incorporated into LOOPER libraries, avoiding its use in sublibraries that contain these problematic codon anticodons will be required. Since the originally reported functionalized LOOPER library uses the 3' dinucleotide to encode the functional group, as in ANNXX, where XX is the encoding dinucleotide, avoiding the use of XX = TT, TA, and AT would solve this issue. Anticodon libraries modified with naphthyl groups, **11c**, were also found to polymerize efficiently by LOOPER with good yield and high fidelity with modest codon bias (**Table 4.1**, entry 5). Only one anticodon sequence, ACGCC, was found to polymerize with an unacceptably high level of error of 35.3%. Noteworthy is that naphthyl-modified ACGCC when linked by HMDA resulted in just 3.8% error, suggesting that the length of the linker can greatly influence the fidelity of certain codons. Notwithstanding, by avoiding napthyl encoding by XX = CC, this high level of error can be avoided. Modification of anticodon libraries with an EDA-linked alcohol group proved to be the most problematic. While overall yield and fidelity were high with modest codon bias (**Table 4.1**, entry 6), several anticodons polymerized with unacceptably high levels of error, including four above 50% error rate: AATTA (52.1%), ATTAT (54.8%), ATTTA (55.3%), and ATTTT (55.7%). This unusually high level of error was also observed with the HMDA linker, albeit to a lesser extent, with three anticodons exhibiting an error rate above 25%. Thus, when encoding for an alcohol

modification, XX = AA, AT, TA, and TT should be avoided. Indeed, we have successfully encoded problematic alcohol modifications with XX = GG with excellent fidelities.[11, 14]

4.4 Conclusions and acknowledgements

In conclusion, we have examined the prospect of decreasing the linker length between the adenosine nucleobase and the displayed functional group during T4 DNA ligase-catalyzed LOOPER. Displaying bulky functional groups directly on the 8-position of dA via an MA linker results in undesired codon bias during LOOPER and high levels of error for large portions of the codon set. We observed that EDA was an ideal linker with respect to LOOPER efficiency, fidelity, and codon bias, and developed an efficient convergent synthesis of modified anticodon libraries linked by EDA using commercially available phosphoramidites. The EDA linker does require adjustment of codon design to minimize codon bias and loss of sequence space during in vitro selections. While this study was intended to address the concern of the long linker length used during LOOPER, this study has also indirectly highlighted the delicate balance of LOOPER. While new insights into the observed limitations of LOOPER are expected to come to light with the recently solved crystal structure of T4 DNA ligase,[22] practitioners of the method are encouraged to conduct duplex DNA sequencing when adjusting chemical modifications, linker length, or codon set sequence beyond reported specifications before pursuing in vitro selections.

Financial support for this work was provided by the National Science Foundation (1506667), the Natural Science and Engineering Research Council of Canada, and the

Ontario Ministry of Research and Innovation. Y. M-A. was supported by an Ontario Graduate Scholarship.

4.5 Experimental details

General Information

Unless otherwise noted, water was purified with ELGA Flex 3 purification system. DNA oligonucleotides without amine modification were purchased from Integrated DNA Technologies. DNA oligonucleotides with amine modification were synthesized on a Bioautomation Mermade 12 synthesizer. All materials and reagents used for oligonucleotide synthesis were purchased from Glen Research. All oligonucleotides were synthesized and deprotected according to the manufacturer's protocols. Oligonucleotides were purified by reverse-phase high-pressure liquid chromatography (HPLC, Agilent 1260) using a C18 stationary phase (Eclipse-XDB C18, 5 μm, 9.4 x 200 mm) and an acetonitrile/100 mM triethylammonium acetate gradient. Oligonucleotide concentrations were quantitated by UV spectroscopy using a Nanodrop ND2000 spectrophotometer. Noncommercial oligonucleotides were characterized by LC/ESI-MS using a C18 column at ambient temperature with a mobile phase of 100% 6 mM triethylammonium carbonate (TEAB) to 80% MeOH and 20% TEAB over 10 min, and a flow rate of 0.3 mL/min. A complete list of all DNA sequences, including templates, barcodes, and adapters, is provided in the supporting information.

Synthesis of amino-modified pentanucleotides

Pentanucleotide anticodons were synthesized using a DMT-ON protocol on a 1 μmol scale (1000 Å CPG column). Bromo-modifier 8'-bromo-dA (Glen Research 10-1007), dA+dC+dG+dT-CE Phosphoramidite (Glen Research 10-1000, 10-1010, 10-1020, 10-1030), Chemical Phosphorylation Reagent II (10-1901) were incorporated as specified by the manufacturer. Following synthesis, the oligonucleotide was cleaved and coupled to ethylene diamine by incubation at 85°C in 440 μL of a 10:1 mixture of 30% ammonium hydroxide and ethylene diamine for 3 h. The cleaved resin was filtered away by filtration. The oligonucleotide was concentrated under reduced pressure using a speedvac. The residue was then taken up into 100 μL of H_2O and purified using reverse-phase HPLC purification using a [10% acetonitrile in 0.1 M TEAA, pH 7] to [80% acetonitrile in 0.1 M TEAA, pH 7] solvent gradient with a column temperature of 45°C. The purified oligonucleotide was then incubated at room temperature in 400 μL of 40% aqueous acetic acid for 1 h to cleave the DMT group, and then lyophilized. The oligonucleotide was incubated in 500 μL 30% ammonium hydroxide at room temperature for 15 minutes to cleave the CPRII linker. Following deprotection, the oligonucleotide was concentrated under reduced pressure using a speedvac. The dried product was dissolved into 100 μL H_2O and subjected to reverse-phase HPLC purification using a [10% acetonitrile in 0.1 M TEAA, pH 7] to [80% acetonitrile in 0.1 M TEAA, pH 7] solvent gradient with a column temperature of 45°C. The purified oligonucleotide was lyophilized, re-dissolved in water, and measured at $\lambda = 260$ nm.

Functionalization of amino-modified pentanucleotides with carboxylic acid derivatives

A mixture of 25 µL carboxylic acid derivatives (100 mM in DMSO), 25 µL *N*-Hydroxysulfosuccinimide sodium salt (sNHS, 100 mM in 1:1 DMSO/H_2O mixture), 5 µL 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride (EDC, 100 mM in DMSO) and 10 µL DMSO was incubated at room temperature for 30 minutes with vortex. Followed by addition of 30 µL of $NaHCO_3$ buffer (pH 8.9, 500 mM in H_2O) and 10 µL amino modified pentanucleotides (1000 µM in H_2O). The mixture was incubated at room temperature for overnight under vortex. The reaction was then spin down to 100 µL and subjected to reverse-phase HPLC using a [10% acetonitrile in 0.1 M TEAA, pH 7] to [80% acetonitrile in 0.1 M TEAA, pH 7] solvent gradient with a column temperature of 45°C. The purified pentanucleotides were dissolved in water, measured at $\lambda = 260$ nm, and characterized with mass spectrometry.

Functionalization of amino-modified pentanucleotides with cyclic anhydride derivatives

A mixture of 25 µL carboxylic anhydride derivatives (100 mM in DMSO), 25 µL $NaHCO_3$ buffer (pH 8.9, 1 M in H_2O), 14.6 µL amino modified pentanucleotides (1000 µM in H_2O), 22.4 µL of H_2O and 163 µL DMSO was incubated at room temperature overnight under vortex. The reaction was then quenched with 50 µL Tris-HCl buffer (pH 8, 500 mM in H_2O) for 1 hour. The reaction mixture was lyophilized down and reconstituted to 100 µL for subjecting to reverse-phase HPLC using a [10% acetonitrile in 0.1 M TEAA, pH 7] to [80% acetonitrile in 0.1M TEAA, pH 7] solvent gradient with a column temperature of 45°C. The purified pentanucleotides were dissolved in water and characterized with mass spectrometry.

LOOPER with modified pentanucleotides

In a PCR tube was added DNA template (15 pmol in 1.5 µL of water), PR5 primer (22.5 pmol in 2.25 µL of water), FAM-PR6 primer (22.5 pmol in 2.25 µL of water) and 5 µL of 4X ligation buffer (40 mM $MgCl_2$, 24% w/v PEG 6000, 40 mM DTT, 264 mM Tris pH 7), 5.5 µL of water and ATP (0.5 nmol in 0.5 µL of water). The reaction mixture was heated to 90 °C for 2 minutes and then cooled down to 25 °C at the rate of 0.1 °C/s. In this PCR tube was then added functionalized pentanucleotides (480 pmol in 1 µL of water; 4 equivalents/codon), BSA (2 µg in 1 µL of water), 400 U of T4 DNA ligase (New England Biolabs, M0202L). The polymerization was performed at 25 °C for 24 h and then purified with MinElute Reaction Cleanup Kit. The crude polymerized material was separated for analysis using denaturing PAGE (10% TBE, 150 V, 55 °C), stained with ethidium bromide,and then imaged by UV illumination.

Adapter duplex synthesis

In a PCR tube was added 15 µL of 100 µM adapter A and 15 µL of 100 µM adapter B, then the tube was heated to 95 °C for 5 minutes and cooled to room temperature over one hour. Then in this PCR tube was added 4 µL NEBuffer2 10× (New England Biolabs, M0212L), 25 U Klenow Fragment (3'→5' exo^-, New England Biolabs, M0212L), 1 µL dNTP Mix (Thermo Scientific, 10 mM each). The extension was performed at 37 °C for 1 hour. The adapter duplex was purified with QIAquick Nucleotide Removal Kit, and then diluted in 30 µL water. In a PCR tube was added 30 µL purified adapter duplex, 5 µl NEBuffer2 10× (New England Biolabs, M0212L), 25 U Klenow Fragment (3'→5' exo^-, New England Biolabs, M0212L), 5 µL dATP (Thermo Scientific, 10 mM), 5 µL H_2O.

This PCR tube was incubated at 37 °C for 1 hour for A-tailing. Then product was purified with QIAquick Nucleotide Removal Kit, and then diluted in 30 μL water.

Adapter ligation

In a PCR tube was added 10 pmol polymerization products, 200 pmol A-tailing adapter duplex, 10 μL NEBNext Quick Ligation Reaction Buffer 5X, 2.5 μL BSA (2 mg/mL in H_2O), 1000 U T4 DNA ligase (New England Biolabs, M0202L), the total volume of reaction was adjusted to 50 μL with H_2O. Then the ligation was performed at 16 °C for 16 hours. The ligated products were then gel purified.

PCR protocols for sequencing

Each purified adapter ligation product was amplified with a different primer from iTrus_D707 to iTrus_D712. In a PCR tube was added 50 attomole purified adapter ligation product in 10 μL H_2O, 1.25 μL 10 μM Primer B, 1.25 μL 10 μM corresponding iTrus_D7XX primer, and 12.5 μL Q5 High-Fidelity 2X Master Mix (New England Biolabs). The tube was then transferred to a preheated themocylcler (98°C). The PCR cycle was started with 10 s of 98°C denature step, followed by 30 s of primer annealing step (annealing temperature was 55°C for the first two cycles, and 71°C for the rest of the cycles), and 30 s of 72°C extension step. The PCR products were then gel purified.

High-Throughput DNA sequencing Protocol

The concentrations of gel purified samples were determined with Kapa library quantification kit for Illumina libraries (KK4845) on Roche LightCycler 480. Paired-end

Illumina sequencing was performed on an Illumina MiSeq system using the kit v2 with 300 cycles (150bp PE sequencing) at the Georgia Genomics Facility, University of Georgia, Athens, GA, USA. LOOPER fidelity was analyzed using the program *analooper*: https://github.com/HiliLab/analooper.

4.6 Supporting results

General information

Unless otherwise noted, water was purified with ELGA Flex 3 purification system. DNA oligonucleotides without amine modification were purchased from Integrated DNA Technologies. DNA oligonucleotides with amine modification were synthesized on a Bioautomation Mermade 12 synthesizer. All materials and reagents used for oligonucleotide synthesis were purchased from Glen Research. All oligonucleotides were synthesized and deprotected according to the manufacturer's protocols. Oligonucleotides were purified by reverse-phase high-pressure liquid chromatography (HPLC, Agilent 1260) using a C18 stationary phase (Eclipse-XDB C18, 5 μm, 9.4 x 200 mm) and an acetonitrile/100 mM triethylammonium acetate gradient. Oligonucleotide concentrations were quantitated by UV spectroscopy using a Nanodrop ND2000 spectrophotometer.

DNA sequences

The DNA sequences below are written from 5'→3'. <Aam>= Amino-modifer C2 dA; <N>=A/T/C/G;

Templates

PANNNN Template /5Phos/GA TTC GCC TGC CGT CGC ANN NNT NNN NTN NNN TNN NNT NNN NTN NNN TNN NNT NNN NTC ACG TGG AGC TCG GAT CCT

Primers

PR5 /5Phos/GG ATC CGA GCT CCA CGT G

PR6 /5Phos/TG CGA CGG CAG GCG AAT CT

FAM-PR6 /5phos/TG CGA CGG CAG GCG AAT CA/36-FAM/

Adapter A AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC T

Adapter B /5phos/ACT GNN NNN NNN NNN NNN AGA TCG GAA GAG CAC ACG TCT GAA CTC CAG TCA C

iTruS_i7_D707 CAA GCA GAA GAC GGC ATA CGA GAT CTG AAG CTG TGA CTG GAG TTC AG

iTruS_i7_D708 CAA GCA GAA GAC GGC ATA CGA GAT TAA TGC GCG TGA CTG GAG TTC AG

iTruS_i7_D709 CAA GCA GAA GAC GGC ATA CGA GAT CGG CTA TGG TGA CTG GAG TTC AG

iTruS_i7_D710 CAA GCA GAA GAC GGC ATA CGA GAT TCC GCG AAG TGA CTG GAG TTC AG

iTruS_i7_D711 CAA GCA GAA GAC GGC ATA CGA GAT TCT CGC GCG TGA CTG GAG TTC AG

iTruS_i7_D712 CAA GCA GAA GAC GGC ATA CGA GAT AGC GAT AGG TGA CTG GAG TTC AG

Primer B AAT GAT ACG GCG ACC ACC GAG

Pentanucleotides

PA-256 /5Phos/<Ama>NNNN

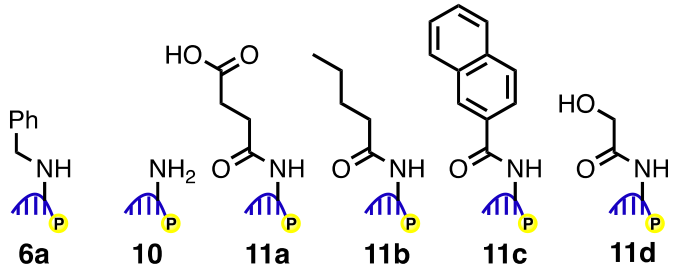


Table S4.1: Mass spectrometric characterization of modified pentanucleotides

Pentanucleotide	Sequence	Calculated mass range	Observed mass range
6a	<P>A*NNNN	1592.31-1752.34	1607.3-1707.3
10	<P>A*NNNN	1545.31-1705.33	1546.0-1707.0
11a	<P>A*NNNN	1645.32-1805.35	1644.2-1807.0
11b	<P>A*NNNN	1629.36-1789.39	1628.4-1790.7
11c	<P>A*NNNN	1699.35-1859.37	1699.9-1860.4
11d	<P>A*NNNN	1603.31-1763.34	1604.4-1760.3

4.7 Reference:

1. Tuerk, C.; Gold, L., Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **1990,** *249* (4968), 505-10.

2. Ellington, A. D.; Szostak, J. W., In vitro selection of RNA molecules that bind specific ligands. *Nature* **1990,** *346* (6287), 818-22.

3. Keefe, A. D.; Pai, S.; Ellington, A., Aptamers as therapeutics. *Nat Rev Drug Discov* **2010,** *9* (7), 537-50.

4. Dunn, M. R.; Jimenez, R. M.; Chaput, J. C., Analysis of aptamer discovery and technology. *Nat Rev Chem* **2017,** *1*, 0076.

5. Zhou, W.; Saran, R.; Liu, J., Metal Sensing by DNA. *Chem Rev* **2017,** *117* (12), 8272-8325.

6. Tan, W.; Donovan, M. J.; Jiang, J., Aptamers from cell-based selection for bioanalytical applications. *Chem Rev* **2013,** *113* (4), 2842-62.

7. Diafa, S.; Hollenstein, M., Generation of Aptamers with an Expanded Chemical Repertoire. *Molecules* **2015,** *20* (9), 16643-71.

8. Kong, D.; Yeung, W.; Hili, R., Generation of Synthetic Copolymer Libraries by Combinatorial Assembly on Nucleic Acid Templates. *ACS Comb Sci* **2016,** *18* (7), 355-70.

9. Wang, Y.; Liu, E.; Lam, C. H.; Perrin, D. M., A densely modified M(2+)-independent DNAzyme that cleaves RNA efficiently with multiple catalytic turnover. *Chem Sci* **2018,** *9* (7), 1813-1821.

10. Guo, C.; Kong, D.; Lei, Y.; Hili, R., Expanding the Chemical Diversity of DNA. *Synlett* **2018,** *29* (11), 1405-1414.

11. Kong, D.; Yeung, W.; Hili, R., In Vitro Selection of Diversely Functionalized Aptamers. *J Am Chem Soc* **2017,** *139* (40), 13977-13980.

12. Lei, Y.; Hili, R., Structure-activity relationships of the ATP cofactor in ligase-catalysed oligonucleotide polymerisations. *Org Biomol Chem* **2017,** *15* (11), 2349-2352.

13. Guo, C.; Hili, R., Fidelity of the DNA Ligase-Catalyzed Scaffolding of Peptide Fragments on Nucleic Acid Polymers. *Bioconjug Chem* **2017,** *28* (2), 314-318.

14. Kong, D.; Lei, Y.; Yeung, W.; Hili, R., Enzymatic Synthesis of Sequence-Defined Synthetic Nucleic Acid Polymers with Diverse Functional Groups. *Angew Chem Int Ed Engl* **2016,** *55* (42), 13164-13168.

15. Lei, Y.; Kong, D.; Hili, R., A High-Fidelity Codon Set for the T4 DNA Ligase-Catalyzed Polymerization of Modified Oligonucleotides. *ACS Comb Sci* **2015,** *17* (12), 716-21.

16. Guo, C.; Watkins, C. P.; Hili, R., Sequence-Defined Scaffolding of Peptides on Nucleic Acid Polymers. *J Am Chem Soc* **2015,** *137* (34), 11191-6.

17. Hili, R.; Niu, J.; Liu, D. R., DNA ligase-mediated translation of DNA into densely functionalized nucleic acid polymers. *J Am Chem Soc* **2013,** *135* (1), 98-101.

18. Chen, Z.; Lichtor, P. A.; Berliner, A. P.; Chen, J. C.; Liu, D. R., Evolution of sequence-defined highly functionalized nucleic acid polymers. *Nat Chem* **2018,** *10* (4), 420-427.

19. Krishnamurthy, V. M.; Semetey, V.; Bracher, P. J.; Shen, N.; Whitesides, G. M., Dependence of effective molarity on linker length for an intramolecular protein-ligand system. *J Am Chem Soc* **2007,** *129* (5), 1312-20.

20. Pfeiffer, F.; Tolle, F.; Rosenthal, M.; Brandle, G. M.; Ewers, J.; Mayer, G., Identification and characterization of nucleobase-modified aptamers by click-SELEX. *Nat Protoc* **2018,** *13* (5), 1153-1180.

21. Gheerardijn, V.; Van den Begin, J.; Madder, A., Versatile synthesis of amino acid functionalized nucleosides via a domino carboxamidation reaction. *Beilstein J Org Chem* **2014,** *10*, 2566-72.

22. Shi, K.; Bohl, T. E.; Park, J.; Zasada, A.; Malik, S.; Banerjee, S.; Tran, V.; Li, N.; Yin, Z.; Kurniawan, F.; Orellana, K.; Aihara, H., T4 DNA ligase structure reveals a prototypical ATP-dependent ligase with a unique mode of sliding clamp interaction. *Nucleic Acids Res* **2018**.

CHAPTER 5

Conclusions and outlook

In this dissertation, we described an alternative approach for introducing various peptide fragments onto a ssDNA scaffold to potentially inhibit protein-protein interactions. Unlike DNA polymerase which has only 4 canonical substrates, DNA ligase enables the recognition of various oligomers. With the increased encoding ability, we proposed ligase-catalyzed oligonucleotide polymerization (LOOPER) in generating various functionalized ssDNA. Optimization of ATP concentration, template architecture, and peptide attachment to the pentanucleotide enabled efficient polymerization of these challenging substrates. Peptides ranging from 2-8 amino acids in length with a wide variety of functionality were found to be within the scope of this method.

We then further demonstrated that the LOOPER process can combinatorially generate ssDNA-scaffolded peptide libraries from co-monomer libraries comprising up to 64 members. The 16-membered codon sets, WSWST and SWSWT, were particularly effective at generating DNA-scaffolded peptide libraries with good efficiencies and at fidelities of up to 99 %. A mock selection was performed with 190-fold enrichment for a single round. This approach should enable the incorporation of 16 different peptides throughout an evolvable ssDNA polymer. Applying this technology to other types of fragments, including carbohydrates or synthetic oligomers would be of great potential. The ability to sequence-specifically incorporate multiple peptide fragments throughout an evolvable ssDNA polymer should enable the in vitro selection of ssDNA-scaffolded

peptides that harness the power of hetero-multi-valency for molecular recognition of protein targets.

In addition, we have proposed two approaches for adjusting the linker length between modification and oligonucleotides. By truncating down to ethylenediamine linker, LOOPER still enables efficient and sequence-specific incorporation. The truncated linker has the potential to facilitate the in vitro selection of diversely-modified DNA aptamers with improved binding thermodynamics.

The further effort could be directed into following parts: designing peptide fragments, functionalization and in vitro evolution towards protein of interest. Previously, the selections toward protein targets were not successful. One of the possible reasons was lacking a rational design for the peptide fragments. For any of the previously described DNA display strategies of peptides, the peptide fragments applied have shown detectable binding towards targets of interest by themselves (micromolar range K_d). To design or identify peptide fragments that present binding toward target of interest would be crucial. Either peptide assay, phage-display or mRNA-display of peptide fragments would enable the identification of peptide fragments in the first place. The identified peptide fragment will then get screened towards binding sites on the protein of interest. After confirming different fragments bind to different sites, the peptide fragments could be determined. In addition, even we have proved up to 8-mer peptides could be tolerated, more complexity might be required. For instance, peptides up to 15-mer and cyclic peptides presented better binding compared to short linear peptides. In that, re-optimization of polymerization will be needed. Modification position on the pentamer might be particularly interesting in this case, as other positions instead of 5'-end were also

tolerated by T4 DNA ligase. After further increasing modification size, they might present differently as well. By expanding the complexity of modification, we also increase the initial peptide fragments library which could generate more hits in return. Finally, with the previous steps, the in vitro evolution towards protein of interest would be more approachable. However, the success of that also relies on the procedure as well. Performing the steps without target of interest helps reveal the potential problems in the system. After solving those problems, perform multiple rounds of selection and monitor the C_T value and T_m value for PCR amplification after each round helps pre-evaluating the selection outcome. The DNA-scaffolded peptide fragments could then be identified towards protein target which potentially enables inhibition for protein-protein interactions.