MULTI-DOMAIN SELECTION OF APTAMERS FOR BACTERIAL PROTEINS

MULTI-DOMAIN SELECTION OF APTAMERS FOR BACTERIAL PROTEINS:

TARGETING *FUSOBACTERIUM NUCLEATUM* DNAK

BY MARIA ALEJANDRA REY RINCON

B.SC.

A Thesis Submitted to the School of Graduate Studies in Partial Fulfilment of the Requirements for the Degree Master of Science

McMaster University

MASTER OF SCIENCE (2020)

Biochemistry & Biomedical Sciences

Hamilton, Ontario

TITLE:

Multi-Domain Selection of Aptamers for Bacterial Proteins: Targeting *Fusobacterium nucleatum* DnaK.

AUTHOR:

Maria Alejandra Rey Rincon, B. Sc. (University of Western Ontario)

SUPERVISOR:

Professor Yingfu Li

NUMBER OF PAGES: viii, 115

ABSTRACT

Aptamers are nucleic acid ligands that bind to a specific target molecule. They are discovered by *in-vitro* selection, whereby binding sequences are selected from a large library of random sequences through iterative affinity steps. Aptamers are used as molecular recognition elements in aptamer-based, as such, creating aptamers with high affinity and specificity to their targets is important to the field. Ligands with two binding sites have been reported to have enhanced binding affinity than ligands with one binding site. To improve the quality of aptamers for downstream applications, multidomain selection is proposed as a new method for selecting aptamers compatible with dimerization. Here, we applied the multidomain selection approach to *Fusobacterium nucleatum* DnaK and produced aptamers that target the N-terminal domain (NTD) and the C-terminal domain (CTD) of DnaK. The top aptamer for DnaK-NTD had a Kd of 59.7 nM, and for DnaK-CTD had a Kd of 202.0 nM. However, the aptamers did not bind to the full-length DnaK and could not be dimerized. Multiple-site binding offers greater flexibility in the design of detection systems, which could provide higher selectivity and sensitivity than aptamers found through standard approaches. Validation of a method to discover aptamers compatible with dimerization would result in the development of a targeted approach to discover high-quality aptamers for bacterial proteins that can be used in bacteria-detection techniques.

ACKNOWLEDGEMENTS

I would like to express my gratitude to Dr. Li for giving me the opportunity of pursuing my degree in his Lab. His guidance and support were invaluable for this work. I would also like to acknowledge my advisors, Dr. Sara Andres, and Dr. Bruno Salena. Their advice and encouragement throughout this project were very reassuring even at the most difficult points. I consider myself to be fortunate that my committee was made of experts in various fields, which allowed me to see my work from different perspectives. I have learned many lessons in science and in life under their tutelage. Additionally, I would like to thank my lab mates who were teachers, collaborators, mentors, students, and friends. You made this ride enjoyable. I will cherish my memories from my time in the Li Lab.

Moreover, I would like to express a special thanks to my wonderful parents to whom I owe my interest in science; they instilled the unwavering curiosity and thirst for knowledge that lead me to this path. Likewise, I would like to thank my husband for being my companion in this journey. Finally, I dedicate this master's thesis to my family and friends whose love and support have driven me to this day.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SUPPLEMENTAL FIGURES

CHAPTER 1:INTRODUCTION

1.1. Historical overview of functional nucleic acids

The discovery of functional nucleic acids changed our understanding of the central dogma of molecular biology. Initially, nucleic acids were only seen as carriers of information; with DNA carrying the instructions for life, and RNA acting as a temporary intermediary. This limited view has slowly changed since the discovery of functional nucleic acids. Pioneering studies discovered that precursor ribosomal RNA (rRNA) can perform cleavage and ligation reactions to self-splice without the presence of enzymes (Kruger et al., 1982). Additionally, it was discovered that proteins like polymerases and ribosomes recognize consensus sequences because of the identifiable three-dimensional structure mRNA adopts at their binding sites (Andrake et al., 1988; Gauss et al., 1987; Y. M. Liang et al., 1988). Thereby, these findings highlight an important relationship between the structure adopted by RNA and recognition by a protein. Indeed, these early findings demonstrated that nucleic acids could carry functional roles that diverge from their canonical functions. Thus, the field of functional nucleic acids was born.

Naturally occurring functional nucleic acids inspired researchers to develop new synthetic forms with innovative functions and characteristics. To this end, the method of *in-vitro* selection of functional nucleic acids was proposed in 1990 by three independent groups (Ellington & Szostak, 1990; Robertson & Joyce, 1990; Tuerk & Gold, 1990). Robertson and Joyce (1990) used *in-vitro* selection to

discover sequences of catalytic RNA (ribozymes) with a novel specific cleavage activity. Here, they took a group I intron from *Tetrahymena thermophila* with RNA splicing activity and selected a variant mutant ribozyme that could cleave a single stranded DNA (ssDNA) substrate. Soon after, deoxyribozymes (DNAzymes) were also discovered to be capable of catalytic activity (Breaker & Joyce, 1994). Since then, this method has been used to isolate Dnazymes for a multitude of functions; such as, DNAzymes able to catalyze ligation of RNA (Flynn-Charlebois et al., 2003), DNA capping (Li et al., 2000), and RNA branching, DNA phosphorylation and cleavage (Wang & Silverman, 2003).

Similarly, *in-vitro* selection was also used to discover nucleic acid ligands that bind to target substrates. Tuerk and Gold (1990) used a randomized library of RNA sequences to isolate sequences able to bind T4 DNA polymerase. In this experiment, they were able to isolate the consensus sequence found in the bacteriophage's mRNA, and a related mutant version. Similarly, Ellington and Szostak (1990) discovered RNA molecules able to specifically bind organic dyes through *in-vitro* selection. Together these two discoveries demonstrated the potential of nucleic acid as ligands against proteins and other organic targets. The nucleic acid ligands selected were named "aptamers" after the Greek word *aptus* meaning "to fit" (Ellington & Szostak, 1990). Since then, aptamers have been discovered for an extensive array of targets, such as proteins and other organic molecules, as well as inorganic molecules.

1.2. Aptamers

Aptamers are nucleic acid ligands made of RNA or ssDNA. They bind targets following the "key in the lock" principle by structurally fitting with their binding partner. Single-stranded nucleic acids fold into three-dimensional structures because their nucleotides tend to interact through π–π stacking, hydrogen bonding, electrostatics and their flexible carbon backbone allows them to adopt many conformations (Gelinas et al., 2016). These structural principles suggest that the structure of aptamers depend on their sequence, and that many different shapes of aptamers may exist. As a result, aptamers have been developed to bind a wide array of molecules with high affinity and specificity (J. F. Lee et al., 2004). Aptamers have become of great interest in the field of molecular detection because of their stability without refrigeration, tolerance to cycles of denaturation and renaturation, ease of synthesis and low price (G. Mayer, 2009). Aptamers are currently being used as effective molecular recognition elements in biosensors. Indeed, aptamers have been recognized as powerful tools for the development of sensors (Dhiman et al., 2017; McConnell, Morrison, et al., 2020; Reid et al., 2020), bioanalytical applications like imaging (Dong et al., 2020; Gao et al., 2018; McConnell, Cozma, et al., 2020), and even in therapeutics (Bazak et al., 2015; Fallah et al., 2019; Srinivasarao & Low, 2017). Therefore, development of aptamers against biomolecules, such as biomarkers of disease and infection, is of great interest in the field.

To develop efficient technologies, it is imperative that aptamers have high affinity and specificity to their target. Aptamers are discovered through the process of *in-vitro* selection or SELEX (systematic evolution of ligands by exponential enrichment) (Tuerk & Gold, 1990) **(Figure 1.1)**. It starts with a library of ssDNA or RNA with 10^9 to 10^{15} unique random sequences. Libraries are composed of a central random region (typically 30 to 50 nucleotides) flanked by primer-binding regions. In principle, a large library of random sequences should contain aptamers able to bind the target of interest. To select for binding sequences, the library is incubated with the target of interest in an incubation step. The target is washed, and the unbound sequences are discarded. The sequences able to bind the target are eluted and amplified with PCR. These sequences are processed and used as the library in subsequent rounds of selection. Thereby, through this iterative affinity screening, binding sequences with high affinity are enriched, and aptamers are selected. Thus, it is crucial to separate the bound and unbound DNA sequences in SELEX; to this end, many SELEX methods have been developed to optimize separation depending on the target and purpose of the selection.

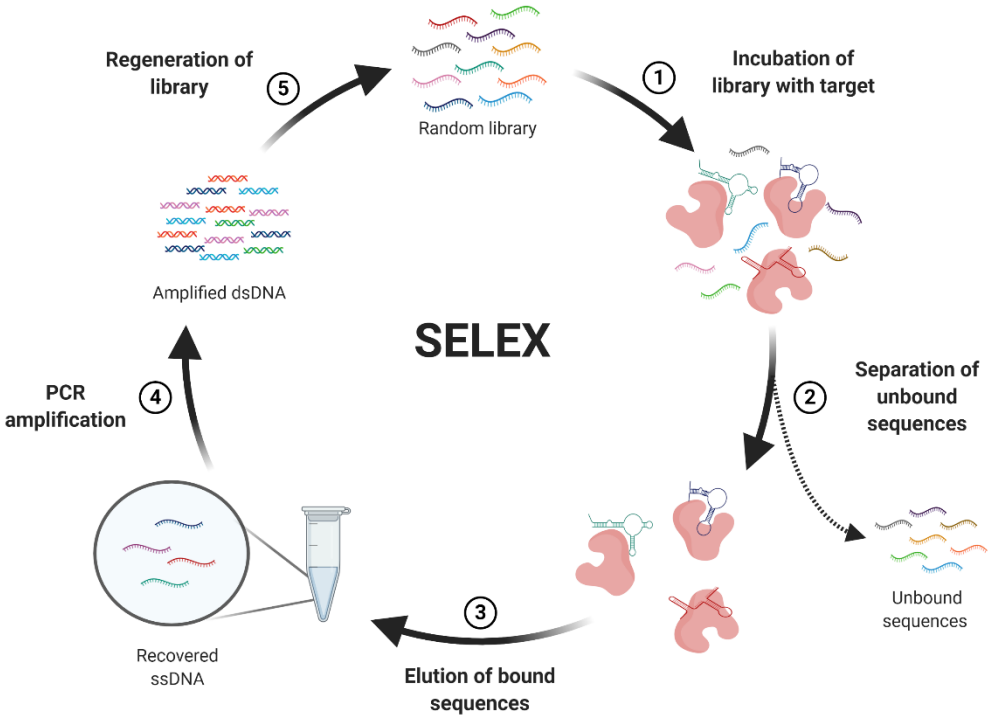


Figure 1.1 Overview of SELEX. The SELEX (systematic evolution of ligands by exponential enrichment) protocol starts by (1) generating a random library of nucleic acids, which is incubated with a target of interest. (2) Fractions of bound and unbound aptamers are separated, and unbound sequences are discarded. (3) Binding sequences are eluted and (4) amplified by PCR. (5) Finally, the library is regenerated and used for a new cycle of selection. Selection is conducted for multiple cycles until binding sequences enrich in the pool of ligands.

1.3. Bivalent ligands and multivalent interactions

Multivalent interactions occur when one given organic molecule has multiple simultaneous molecular associations with its target ligand or ligands (Mammen et al., 1998). Multivalent interactions are widespread in biology and are common when precise molecular recognition is required (Mammen et al., 1998). Two potential ligand binding-sites increase the binding-site's local concentration and the probability of an interaction (Vauquelin & Charlton, 2013). This principle is known as avidity, and it describes the apparent affinity of a multivalent interaction (Mammen et al., 1998). Similarly, the target ligands experience forced proximity upon dissociation because a free binding-site is more likely to be nearby and available for rebinding, which results in a reduced rate of dissociation (Vauquelin & Charlton, 2013). These principles hold true for all bivalent ligands including homo-bivalent ligands with one binding site, or hetero-bivalent ligands that have two different binding sites (Vauquelin & Charlton, 2013). As a result, developing ligands that exploit the properties of multivalent interactions has been of interest in the field of biochemical engineering.

Not surprisingly, bivalent aptamers have been studied in the past. Bivalent aptamers are made when two aptamers are joined through a linker without losing their structural properties; thus, each side of the aptamer folds independently and binds its respective epitope. In theory, bivalent aptamers should have improved binding affinity due to the kinetics of multivalent interactions. Thrombin binding aptamers (TBAs) have been widely studied for the past 25 years (Bock et al.,

1992); Thus, many aptamer moieties have been developed and extensively characterized. High affinity TBAs are of interest because they can inhibit thrombin and act as anticoagulants (Bock et al., 1992). Hetero-bivalent TBAs have been developed by multiple groups, which have consistently reported increased affinity and improved anticoagulant activity (Ahmad et al., 2012; Kim et al., 2008; J. Müller et al., 2008; Jens Müller et al., 2007; Tian & Heyduk, 2009). Similarly, hetero-bivalent and homo-bivalent aptamers for vascular endothelial growth factor (VEGF) have been reported to have decreased apparent Kd values compared to their monovalent counterparts (Hasegawa et al., 2008; Nonaka et al., 2010). Enhanced avidity was reported for the dimerized aptamer compared to the monovalent aptamer for heat shock factor 1 (HSF1) (Zhao et al., 2013). Overall, these studies demonstrate the potential of bivalent interactions to increase ligand binding affinity.

However, the success of bivalent aptamers is dependent on the structural conditions for the interactions. For example, the HSF1 homo-bivalent aptamer only works because HSF1 exists as a homotrimer (Zhao et al., 2013); thus, this kind of aptamer arrangement can only be used in multimeric proteins. For monomeric or heteromeric proteins, aptamers that bind different sites are needed. In these instances, the aptamers must recognize distinct sites in the protein and their binding should be compatible. That is, they must bind the protein without interfering with the other. Indeed, it has been reported that adapting existing aptamers into bivalent forms can often lead to no increase in affinity due to

incompatibility of the binding sites (Manochehry et al., 2019). While it is possible to develop aptamers that bind two distinct sites on the target from one selection process, this approach would depend on random chance. The selected aptamers may still be competing for a binding site or be incompatible with dimerization. Therefore, a site-targeted selection method for aptamer discovery should be used to find aptamers that bind distinct sites of the protein and that, as a result, can be readily used for building multivalent aptamers.

1.4. Bead-based SELEX

In bead-based SELEX the target molecule is immobilized on the surface of beads. This version of SELEX was originally introduced by conducting selection on chloroaromatics that were chemically linked onto magnetic beads (Bruno, 1997). This method uses beads made of solid resins to immobilize the target molecule permanently through chemical modifications of the target, or transiently through non-covalent interactions. Common covalently modified resins include agarose and Sepharose. Alternatively, streptavidin, Nickel-nitrilotriacetic acid (Ni-NTA) and glutathione resins are commonly used for non-covalent immobilization of protein targets. Therefore, bead-based selection provides a simple platform to easily immobilize recombinant protein targets since most recombinant proteins are tagged with histidine or glutathione S-transferase (GST) for purification. Indeed, this method of selection is so convenient that it has become one of the most used methods for selection of aptamers against protein targets. Once the target is immobilized it can be easily and effectively handled to conduct selection.

Here, the immobilized target is incubated with the library and separation of bound and unbound oligonucleotides can be done by aggregating the beads with centrifugation or with a magnet for magnetic beads. Enabling efficient separation of the bound DNA from the unbound fractions. Moreover, the loaded beads can be readily processed, which allows for the design of flexible experiments with multiple stages of positive and negative selection.

Creative experimental designs have been previously used in bead-based selection to selectively target a specific aptamer binding site. Targeting selections to a specific protein epitope with bead-based SELEX was pioneered by Gong *et al.* (2012) with the multivalent aptamer isolation (MAI-SELEX) strategy. MAI-SELEX aims to select for aptamer pairs that bind different sites of the target protein such that they can be used in tandem for downstream applications (Gong et al., 2012). In this approach, bead-based SELEX was conducted on αVβ3 integrin for five rounds; then, the selected RNA pools were spilt into αV and β3 binders by performing one round of selection with a decoy protein (αIIβ3), which sequesters β3 domain binding aptamers, followed by selection with αVβ3, which pulls the αV binding sequences. Effectively, the aptamer pool was split to yield two domain-specific aptamers through a creative experimental design. Nonetheless, this approach is limited to homologous proteins that share common subdomains which severely limits the cases when it can be used.

More recently, sophisticated site-targeting approaches have been proposed by other groups. Tucker *et al.* (2018) proposed a method to target a specific protein

sites of the HECT domain of the large protein nuclear WW domain containing E3 ubiquitin ligase 1 (WWP1). To pin-point this specific binding site, they conducted conventional selection on the HECT domain for eight rounds, then, they branched the selection against the C-lobe, N-lobe, and the whole domain (Tucker et al., 2018). While this approach could be used to target selection towards a distinct site of the protein, they were unable to isolate aptamers for the N-lobe, and the C-lobe aptamers had poor affinity (low micromolar). Thus, their method was unable to produce aptamers targeting different domains. An even more discerning site-specific targeting method was proposed by Sedighian *et al.* (2018). Their approach, staggered-SELEX, aimed to target selection towards two specific surface peptides of the Staphylococcal enterotoxin type A (SEA) that are only five and six amino acids long (peptide 1 and 2, respectively). The selected aptamers needed to have a precise binding site because SEA is highly homologous to the Staphylococcal enterotoxin types B, C, and D (SEB, SEC and SED); thus, two SEA specific surface peptides were identified for precision targeting to avoid cross reactive aptamers. Selection was done in two stages in staggered-SELEX. First, conventional bead-SELEX for four rounds with SEA as the positive target and SEB, SEC, and SED as negative targets. Then, five rounds of selection were conducted where the DNA pool was incubated with peptide 1, bound sequences were eluted and consecutively used against peptide 2. Hence, staggering the positive targets. Aptamers with high affinity towards SEA and no cross reactivity against SEB, SEC and SED were identified using this method (Sedighian et al.,

2018). Undeniably, this is an extreme example of site-targeting a selection. However, this strategy also relies on counterselection strategies with homologous proteins with high sequence similarity. As such, it can only be used in special cases.

The examples above show the flexibility and versatility of selection strategies that used bead-based SELEX protocols. They also demonstrate that site-targeting in selection is possible with careful experimental design. As a result, an experimental approach can be designed using bead-based selection to produce site-specific aptamer pairs that can be used to construct bivalent aptamers.

1.5. Multiple domain selection

The Li lab has started a new approach for aptamer selection called multidomain selection (MDS) to generate aptamers that are compatible with multimerization. Then, through this approach, high affinity bivalent aptamers can be developed. To overcome the limitations discussed in previous sections, the selection method must pin-point the binding sites of the aptamers towards independent regions of the protein. Thus, in MDS the target protein is divided into two domains, then, selection is performed on each individual domain **(Figure 1.2)**. This method allows selection to be conducted in a targeted manner; thus, it is hypothesized that the aptamers selected with MDS can overcome the structural limitations often faced by aptamers found through regular selection. That is, the aptamer binding sites are distinct and likely to be distant from each other, then, the domain-specific aptamers are more likely to be compatible when assembling a bivalent aptamer.

Moreover, the protein targets can be chosen to have desired characteristics, like high cellular abundance and structural stability. MDS aims to exploit the binding kinetics of bivalent ligands; that binding a target with two sites increases affinity by increasing the number of binding contacts, the binding sites' local concentration, and the time the target is bound (Vauquelin & Charlton, 2013). Therefore, it is hypothesized that bivalent aptamers will have higher affinity to the target than each individual aptamer, and that two separate binding sites will make bivalent aptamers more specific to the target.

To test the MDS approach, bacterial DnaK was chosen as the protein target. DnaK is part of the heat shock protein 70 (Hsp70) protein family; it is an endogenous bacterial protein that is highly expressed in the cell and it is also highly conserved across species. DnaK has two functional domains the N-terminal domain (NTD) holds a nucleotide binding domain with ATPase activity, and the C-terminal domain (CTD) which is the substrate binding domain that binds to unfolded polypeptides. A short flexible linker region serves as a structural and functional connection between the two domains (Matthias P. Mayer & Gierasch, 2019). DnaK has been previously found to be functional when split into its individual domains (Bertelsen et al., 2009; Buchberger et al., 1995). Moreover, as a chaperone protein, DnaK is highly stable under a wide range of conditions. As such, DnaK is a suitable candidate for MDS.

The Li lab has used MDS to develop a high affinity bivalent aptamer in work done by a previous student (Arrabi, 2017). DnaK from the infectious bacteria

Clostridium difficile was chosen as the protein target. The anaerobic bacterium *C. difficile* was targeted because it causes *Clostridium difficile* infection (CDI) in humans. CDI is an intestinal infection that can present mild to severe symptoms. *C. difficile* DnaK DnaK-CTD, and DnaK-NTD were purified and aptamer selection was performed on each protein. Viable aptamers were produced for all targets. The top aptamer for the full-length DnaK had a dissociation constant (Kd) of 160 nM and binds the protein at the N-terminal domain, while the aptamers selected from DnaK-NTD and DnaK-CTD had a Kd of 206 nM and 69.5 nM, respectively. The top aptamers selected from the domains were joined using a variety of linkers and binding was characterized; the ultimate bivalent aptamer had a Kd of 1.66 nM. Thus, the bivalent aptamer had a much higher affinity than any individual aptamer. The bivalent aptamer was also tested against four different species of bacteria and found to be specific to *C. difficile*, except in the presence of very high concentrations of *Escherichia coli (E. coli)*. Overall, this work showed the power of creating bivalent aptamers for improving aptamer affinity.

However, it is still inconclusive whether this method can be repeated to consistently produce high quality bivalent aptamers that outrank those produced from single selection. This work also leaves the open question of whether the increase in affinity of the aptamer will correlate to the aptamer specificity. Certainly, increased affinity may be for nought if the bivalent aptamers produced for DnaK cannot differentiate between bacterial species. Thus, it should be

examined whether MDS can produce aptamers that can distinguish between bacteria species and even strains. Indeed, the MDS method for aptamer development needs to be challenged and further validated.

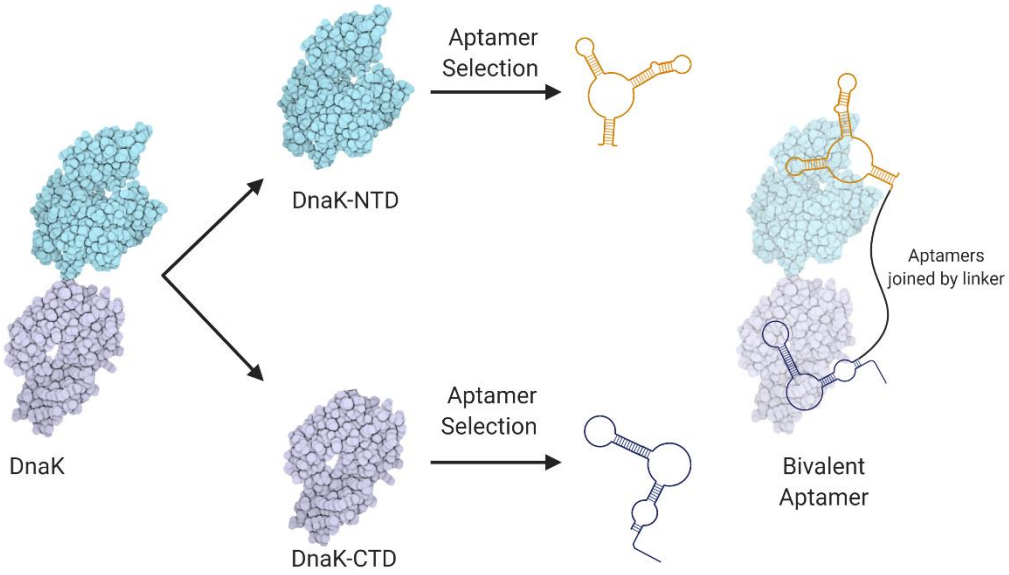


Figure 1.2 Overview of multidomain selection (MDS) approach. The target protein (DnaK) is split into two domains: DnaK-NTD and DnaK-CTD. Each individual domain undergoes aptamer selection to produce aptamer candidates that bind distinct sites. Aptamers are joined by a linker to produce a bivalent aptamer.

1.6. Colorectal cancer and *Fusobacterium nucleatum*

Fusobacterium nucleatum is a bacterial species from the *Bacteroidaceae* family found in human oral and gut microbiota. It is an anaerobic, non-spore-forming, and gram negative bacterium (Bolstad et al., 1996). *F. nucleatum* behaves as an opportunistic pathogen that thrives under inflammatory conditions in the oral and intestinal epithelial cells; as a result it is associated with periodontal disease, oral cancers and colorectal cancer (CRC) (Gholizadeh et al., 2016). *F. nucleatum* has been reported to works synergistically with *Porphyromonas gingivalis* by promoting the inflammation involved in the development of periodontal disease, and that drives oral tumorigenesis (Brennan & Garrett, 2019). Similarly, *F. nucleatum* has been associated with other types of cancer like pancreatic cancer and cervical cancer (Gaiser et al., 2019; Kyrgiou et al., 2017). Therefore, detection of *F. nucleatum* has become of great interest for diagnosis of disease.

In 2012, pioneering research on the relationship between the gut microbiome and CRC found that *F. nucleatum* is more abundant in the colon of patients with CRC than healthy controls (Castellarin et al., 2012; Kostic et al., 2012). *F. nucleatum* was also found to be abundant in the colon of patients with pre-cancerous adenomas (McCoy et al., 2013). A growing number of studies have linked CRC with *F. nucleatum* in diverse cohorts of populations from across the world; such as European, East Asian, Latin American and North American populations (Dai et al., 2018; de Carvalho et al., 2019; Guo et al., 2018; Lauka et al., 2019; D.-W. Lee et al., 2018; Liu et al., 2020; Mima et al., 2016; Mizutani et al., 2020; Russo et al.,

2018; Tunsjø et al., 2019; Vogtmann et al., 2016; Yachida et al., 2019; Yang et al., 2019). Abundance of *F. nucleatum* has also been found to increase with cancer progression and to negatively correlate with patient survival (Flanagan et al., 2014; Lauka et al., 2019; Mima et al., 2016; Wei et al., 2016). This enrichment of *F. nucleatum* was also detectable in the faecal samples of colorectal cancer patients (Q. Liang et al., 2017; Yu et al., 2017; Zackular et al., 2014). As a result, *F. nucleatum* has become a potential biomarker for CRC.

In recent years, the potential of *F. nucleatum* as a biomarker for CRC screening has been explored. First, it was reported that coupling fecal occult blood tests (FOBT) with the detection of *F. nucleatum* improves accuracy of CRC detection (Zeller et al., 2014). Later on, it was also reported to increase the accuracy and sensitivity of faecal immunochemical test (FIT) screening of CRC (Q. Liang et al., 2017; Wong et al., 2017). Furthermore, *F. nucleatum* has been consistently found to be a microbial biomarker for CRC present in faecal samples by subsequent studies (Grobbee et al., 2019; Yu et al., 2017; Zhang et al., 2019). Thus, making a simple and reliable detection method for *F. nucleatum* could provide a simple method for the screening and detection of CRC.

1.7. Thesis objectives

The objective of this thesis is to examine the MDS approach and challenge its effectiveness. To do so, DnaK of *Fusobacterium nucleatum* has been targeted for multidomain aptamer selection using bead-based SELEX. This work aims to test the ability of MDS to effectively produce aptamers against the domains of DnaK that can be used to create a bivalent aptamer, and to characterize the binding properties of these aptamers. It is hypothesized that MDS can be used to produce bivalent aptamers with higher affinity to the target than each individual aptamer. To do this, the following experimental objectives were proposed.

1) Design plasmids for *F. nucleatum dnak, dnak-ctd* and *dnak-ntd* and clone them into a protein expression system.
2) Express and isolate the protein targets with high purity.
3) Conduct selection on the N-terminal of DnaK (DnaK-NTD) and the C-terminal of DnaK (DnaK- CTD).
4) Identify aptamer candidates and characterize the binding properties to their respective domains.
5) Characterize the binding of aptamers towards the full DnaK.
6) Design a bivalent aptamer from sequences of single domain selections, and characterize the bivalent aptamer compared to the single aptamers.

CHAPTER 2: MATERIALS AND METHODS

2.1. Designing an expression system for *Fusobacterium nucleatum* proteins DnaK, DnaK-CTD and DnaK-NTD

Expression system

To optimize protein expression, the pET15b expression vector was chosen. pET15b is under the strong bacteriophage T7 promoter and is expressed with the activation of the Lac operator. Furthermore, the pET15b produces His-tagged proteins that can be purified with nickel column system. The gene synthesis service from Biomatik was used to synthesize *F. nucleatum dnak*, which also included codon optimization for expression in *E. coli*. The *dnak*, and the *dnak-ctd* and *dnak-ntd* constructs were cloned into the pET15b vector. Biomatik also provided a quality control report that ensures the sequence, purity, and yield of the vectors (5 µg). Plasmids were resuspended and stored according to the manufacturer's instructions.

Transformation of electrocompetent cells

Plasmids containing *dnak*, *dnak-ctd* and *dnak-ntd* were transformed into *E. coli* DH5α cells for storage and proliferation of the plasmid, and into *E. coli* BL21 cells for protein expression. Electrocompetent *E. coli* cell stocks were mixed with each plasmid (50 ng) and applied to cuvettes for electroporation. Cells were electroporated twice at 1650 volts, constant current. Cuvettes were washed with 700 µL of LB (25 g/L) to retrieve the cells, which recovered for one hour at 37 ºC

(250 RPM). Cells (100 µL) were plated on LB-Amp (100 µg/mL) agar plates and incubated overnight (37 ºC). Transformed cells formed distinct round colonies on the plate, the number of colonies on the plate varied with the efficiency of electroporation and the recovery time.

Plasmid extraction and sequencing

To verify the gene sequences in the plasmids, *E. coli* DH5α transformed with *dnak, dnak-ctd* and *dnak-ntd* were cultured to extract the plasmids for sequencing. Liquid LB-Amp media (100 µg/mL,100 mL) was inoculated by single colonies and grown overnight at 37 ºC. These cultures were also used to make glycerol stocks for plasmid storage. Plasmids were extracted using the PureYield™ Plasmid Miniprep System (Promega) following the manufacturer's instructions. The concentrations of the extracted plasmids were measured using NanoVue Plus spectrometer (GE Healthcare). Plasmid minipreps for *dnak* (77 ng/µL), *dnak-ctd* (67 ng/µL) and *dnak-ntd* (75 ng/µL) were sent for sequencing. Genes were sequenced from the T7 promoter and the T7 terminator; the sequencing files from each direction were consolidated into one gene sequence.

2.2. Expression and purification of target proteins: DnaK, DnaK-NTD and DnaK-CTD

Preparation of reagents

Reagents were prepared in large quantities for large scale protein purification. LB broth (25 g) was dissolved in 1 L of water in a 4 L Erlenmeyer flask, and

autoclaved. After cooling, 1 mL of Ampicillin (100 g/L) was added to each flask. LB was stored at 4 °C. Buffers for the protein purification process were prepared using autoclaved water and solutions for NaCl (5 M) and TRIS-HCl (1M, pH 8.0). Binding buffers (A) and an elution buffers (B) were prepared for both the Ni-NTA affinity column (NiA, NiB) and the anion exchange column (QA, QB) as follows: NiA (TRIS-HCl pH 8.0 50 mM, NaCl 500 mM, imidazole 20 mM, β-mercaptoethanol 5 mM, 5% glycerol), NiB (TRIS-HCl pH 8.0 50 mM, NaCl 500 mM, imidazole 300 mM, β-mercaptoethanol 5 mM, 5% glycerol), QA (TRIS-HCl pH 8.0 20 mM, β-mercaptoethanol 5 mM, 5% glycerol), and QB (TRIS-HCl pH 8.0 20 mM, NaCl 500 mM, β-mercaptoethanol 5 mM, 5% glycerol). Buffers were filtered through a 0. 22 µm filter using a vacuum pump, and stored at 4 °C.

Expression of target proteins in E. coli

Protein expression was achieved by culturing *E. coli* BL21 cells with pET15b vectors with gene inserts *dnak, dnak-ctd* or *dnak-ntd*. Cells were transformed and grown overnight on plates as explained above. Single colonies were used to inoculate 100 mL liquid LB- Amp media (100 µg/mL). Single-colony liquid cultures were incubated overnight; then, they were sub-cultured into 1 L of fresh LB-Amp (1/50) and grown to mid-log phase (OD_{600} ~0.7) Protein expression was induced by adding 1 mL of isopropyl β- d-1-thiogalactopyranoside (IPTG, 1 M) and incubating for 4 hours. All liquid cultures were incubated at 37 ºC in shaker (250 RPM). Each one litre of cell culture was pelleted by centrifugation (15 min, 4000 x g), resuspended in 25 mL of PBS transferred into 50 mL tube, and re-

pelleted (10 min, 4000 x g). The supernatant was removed, and cell pellets were frozen at -80 ºC for storage.

Purification of proteins AKTA-Start system

Cell pellets were thawed and resuspended in NiA buffer and a mixture of protease inhibitors was added. The cell mixtures were lysed using sonication; the probe was placed into cell mixture, and sonication was done five times for 30 s with 30 s rest intervals. Cell lysates were centrifuged (18,000 x g, 45 min) to separate the cellular debris and the supernatant containing the proteins of interest. Lysates were filtered with 0.22 µm syringe filter.

The AKTA-Start protein purification system (GE Healthcare) was used for two steps of purification. First, the HisTrap HP Ni-NTA affinity column (GE Healthcare) was used to capture His-tagged proteins. Then, the anion exchange column, HiTrap Q HP (GE Healthcare), was used to further separate proteins based on their surface charge. For both columns, a protocol was set up following the manufacturer's instructions. These protocols had a system equilibration step with the binding buffers (5 column volumes) to prime the equipment. Then, the sample was loaded onto the column. Unbound proteins were washed off with a buffer mix containing 5 % elution buffer for 5 column volumes. Finally, bound proteins were eluted using a gradient with increasing concentration of elution buffer from 5 to 100 % (NiB or QB) over 20 column volumes.

Filtered cellular lysates were used for the Ni-NTA affinity column. Bound proteins were eluted from the column with an imidazole gradient (30 to 300 mM) and elution fractions were collected. Fractions containing the protein were detected by UV 280 nm, and these fractions were checked for the presence of the protein of interest. Proteins of interest were identified by running the samples on a denaturing sodium dodecyl sulfate polyacrylamide gel (SDS-PAGE, 12%), which separates proteins by molecular weight, alongside a protein ladder. Gels were stained with Coomassie blue to visualize proteins; the detection of the protein of interest was stablished by size. Samples containing the protein of interest were pooled, diluted in QA buffer (4/5 QA), and further purified through the anion exchange column. Proteins were eluted with a NaCl gradient (25 to 500 mM) and fractions were collected. Fractions containing the protein of interest were identified and further investigated by 12% SDS-PAGE. Fractions containing the proteins of interest at high purity were concentrated using Amicon Ultra centrifugal filter (EMD Millipore). All proteins were concentrated to ~150 µL in selection buffer (SB, HEPES pH 7 50 mM, NaCl 100 mM, KCl 10 mM, $MgCl_2$ 5 mM, 0.03 % Tween-20) with 10% glycerol and stored at - 80ºC in 10 µL aliquots.

Quantification of protein in stock

The concentration of proteins in the stocks was calculated using a Bradford assay. Six bovine serum albumin (BSA) protein standards of varying concentrations (1.5 mg/mL to 0.25 mg/mL) and the protein samples were incubated with Bradford reagent. Then, their absorbance was measured at 595 nm wavelength. A standard

curve was created by plotting the net absorbance of the BSA standards over their concentrations, and the equation of the line was extrapolated. The concentration of the protein samples was calculated using this method and confirmed by running standards and samples on 12% SDS-PAGE with Coomassie staining.

2.3. Overview of protocol for aptamer selection: Magnetic bead SELEX

Immobilization of proteins on magnetic beads

HisPur™ Ni-NTA Magnetic Beads were purchased from Thermo to immobilize the proteins of interest. Manufacturer's specification suggested that the capacity of the beads is 18.75 pmol/µL of bead. Thus, this was used as the theoretical maximum capacity of the bead. Beads were loaded with twice as much protein to completely coat the beads with target protein. For instance, 10 µL of beads would be incubated with 375 pmol of protein to achieve 187.5 pmol of protein loaded. To load the protein onto the beads, beads were diluted in selection buffer (1/20), then pelleted with a magnetic rack to remove the supernatant. Beads were washed three times with resuspension: resuspend beads in selection buffer (300 µL), mix by pipetting, pellet beads with magnetic rack, remove supernatant, and add fresh supernatant. Beads were then resuspended in 200 - 400 µL of selection buffer with the appropriate amount of target protein, the mixture was incubated on a rotating shaker at 4 °C for at least 30 min. The immobilized protein was pelleted and supernatant with excess free protein was removed. Beads were washed three times with resuspension as explained above. Then, beads were rinsed three times:

beads are pelleted, selection buffer (300 µL) is added and pipetted up and down twice without disturbing the beads. Finally, bead-immobilized protein was resuspended to a final concentration of 2 µM. Note that reagents were kept cold on ice throughout the protocol to avoid protein unfolding.

Aptamer selection protocol

Aptamer selection was performed using ssDNA libraries with a 40N random region flanked by known primer sequences. Random libraries were ordered from IDT and purified to remove DNA artifacts from the synthesis process. Sequences were separated in a thick (1.5 mm) denaturing polyacrylamide gels (dPAGE) (10 % polyacrylamide in 1X TBE and 7 M urea), library sequences were excised from the gel and DNA was extracted by the gel-crush and soak method. DNA was recovered by ethanol precipitation, resuspended in water, and quantified with NanoVue. All aptamer selections were started with 1000 pmol of ssDNA library. Selection was conducted in selection buffer. Libraries were denatured by heating at 90 °C for 5 min, then, the ssDNA was allowed to refold by cooling on ice for 10 min, and equilibrating at room temperature for 15 min. Bead-immobilized protein and library were incubated on rotating shaker for 30 min at room temperature. A magnetic rack was used to pellet the magnetic bead-immobilized protein and bound DNA. Beads were washed at least three times with resuspension and rinsed three times as explained above. Protein was eluted from beads with imidazole elution buffer (selection buffer with 300 mM imidazole) (15 min in shaker). Beads were pelleted and the supernatant containing eluted target-

bound DNA was added to 1000 µL of cold 100% ethanol for DNA precipitation. Recovered DNA was resuspended in ~ 20 µL of milliQ water. Counter selection was introduced in later rounds by incubating the DNA library with Ni-NTA beads for 20 min to remove bead-binding sequences, the supernatant with free library was then used for positive selection. Furthermore, 15 µg of free BSA were added to the positive selection to remove any non-specific protein-binding sequences.

Two amplification steps were used during selection. The first amplification step used primers complementary to the ssDNA library, this first round of amplification was termed PCR 1. The second amplification step (PCR 2) used a reverse primer with an overhang, which made the reverse strand heavier than the forward strand. Thus, the strands could be separated by dPAGE. PCR reactions were set up with 1 µM of forward and reverse primers, 200 µM dNTPs, 2 µL of template DNA, 0.5 µL of Taq polymerase in 1x Taq reaction buffer (50 µL reactions). To optimize the number of cycles needed for optimal amplification, PCR tests were conducted. One PCR reaction was set up using 2 µL of eluted ssDNA library as template. The reaction was amplified for eight cycles and 5 µL of sample were removed, then, amplification continued for two more cycles and another 5 µL sample was taken. This process was repeated such that samples were obtained for 8, 10, 12, and 14 cycles of amplification. Samples were run on a 2% agarose gel for 30 min at 160 volts. The gels were imaged with Typhoon and cycle number was chosen for optimal amplification.

Recovered ssDNA library was used as template for PCR 1. A PCR 1 test was conducted as explained above to choose optimal number of cycles. Then, PCR 1 was conducted for 10 reactions. PCR 1 was pooled and stored at -20 °C to maintain a repository of DNA tracking selection. PCR 1 reaction was diluted 1/20 in water, which was used as template for PCR 2 (2 µL). To verify quality of DNA amplification in PCR 1, a PCR 2 test was conducted as explained above. Twenty-eight PCR 2 reactions were conducted, and usually amplified for 10 cycles. The dsDNA of PCR 2 was recovered by ethanol precipitation. DNA pellets were resuspended in 80 µL of water by heating at 90 °C and vortexing. Then, 2X urea loading buffer (urea 1.1 g/mL in sucrose solution: 1X TBE, sucrose 200 g/L, bromophenol blue 0.25 g/L, xylene cyanole FF 0.25 g/L, 0.1 % SDS) was added, and samples were heated at 90 °C for 7 min. Samples were loaded onto a thin (0.75 mm) 10 % dPAGE, and ran at 36 mA for 1 h and 20 min. The gel was wrapped and exposed to UV light where DNA can be visualized as a shadow. The forward and reverse strands run separately on the dPAGE, thus, the shorter forward strand was excised and purified by gel crush and soak method. The ssDNA was recovered by ethanol precipitation, resuspended in 50 µL of water, and quantified by NanoVue. A schematic of magnetic bead SELEX is provided in **Figure 2.1**.

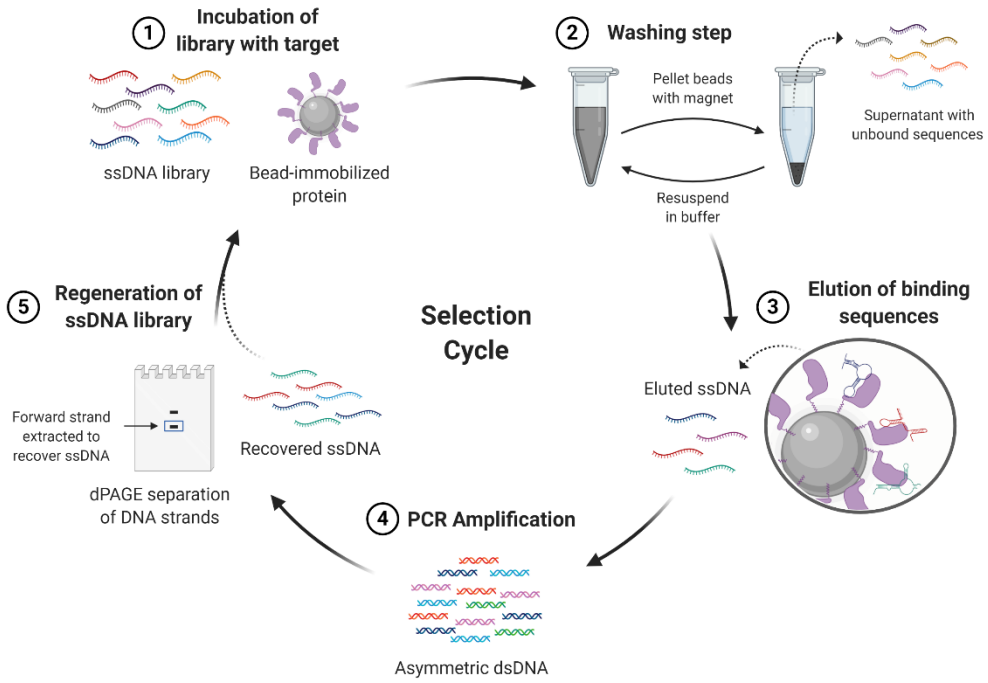


Figure 2.1 Magnetic bead SELEX protocol. The selection protocol was conducted with a random library of ssDNA (82 nt, 40N) and protein immobilized on magnetic Ni-NTA beads. (1) The library and target were incubated at RT on a rotating shaker for 30 minutes. (2) Magnetic beads were washed by pelleting the beads with a magnetic rack, removing the supernatant containing unbound sequences, and resuspending the beads in fresh buffer. The wash cycle was repeated at least three times with increased washing in later rounds of selection. (3) Binding sequences were eluted with elution buffer (300 mM imidazole) and recovered with ethanol precipitation. (4) Eluted DNA was amplified with PCR 1, and with a heavy reverse primer for PCR 2. (5) DNA strands were separated in a 10% dPAGE and the lighter forward strand was extracted to recover the ssDNA library. The selection cycle was conducted for eight to ten rounds.

2.4. Selection of aptamers for DnaK N-terminal and C-terminal domains

Aptamer selection for DnaK-NTD 1 (FN40R-6 Library)

An ssDNA library with 40-nucleotide random region was used for the DnaK-NTD aptamer selection (5’ – AGT GTC AGC CAG TAT AAC CCA – N40 – TAA CCC GGA GTG AAC ACC TA – 3’). The forward primer used in both PCR 1 and PCR 2 was 5' – AGT GTC AGC CAG TAT AAC CCA – 3', for PCR 1 the reverse primer was 5' – TAG GTG TTC ACT CCG GGT TA – 3’, and for PCR 2 the reverse primer was 5' – TTTTTTTTTTTTTTT /iSp18/TAG GTG TTC ACT CCG GGT TA – 3'. Selection started with 1000 pmol of library, for each subsequent round ~100 pmol of library was used. All incubation steps contained 130 nM of DnaK-NTD immobilized on Ni-NTA magnetic beads in selection buffer (SB, HEPES pH 7 50 mM, NaCl 100 mM, KCl 10 mM, $MgCl_2$ 5 mM, 0.03 % Tween-20). Beads were washed three times to remove unbound DNA. A total of ten rounds of selection were completed. The last five rounds included a counter selection step against unloaded Ni-NTA magnetic beads, and BSA was added to the positive selection as a non-specific protein target. The sequence composition of libraries at round seven to ten (N7 to N10), and N2, N4, and N6 were analysed with deep sequencing.

Aptamer selection for DnaK-CTD (FN40R-7 Library)

An ssDNA library was used for the selection against DnaK-CTD (5' –CCA AAC AGC CAG TAT GTC AGT – N40 –TGG GTT ATA CTG GCT GAC ACT – 3'.)

The forward primer used was 5' – CCA AAC AGC CAG TAT GTC AGT – 3', the PCR 1 reverse primer was 5' – AGT GTC AGC CAG TAT AAC CCA – 3', and the PCR 2 reverse primer was 5' – TTTTTTTTTTTTTTT /iSp18/AGT GTC AGC CAG TAT AAC CCA – 3'. DNA sequences were ordered from IDT, purified separately with electrophoresis (thick dPAGE, 10 %), resuspended in freshly autoclaved water, and stored at -20 °C in separate boxes (library and primers). Surfaces were cleaned with 10 % bleach before and after each selection round, and pipetting was done with filter tips only. Tubes containing template were carefully handled and stored separately from primers and other reagents. To begin selection, 1000 pmol of library was used. Each subsequent round of selection used 100 pmol of library. All rounds of selection contained 130 nM of DnaK-CTD and library in selection buffer (SB, HEPES pH 7 50 mM, NaCl 100 mM, KCl 10 mM, $MgCl_2$ 5 mM, 0.03 % Tween-20). For the first five rounds the target was washed three times with resuspension and rinsed three times. Stringency of positive selection was strengthened by increasing washing and rinsing progressively to 10 washes and 10 rinses, and by decreasing incubation time to 20 minutes. Counter selection was conducted after round five. Counter-selection stringency was increased on subsequent rounds by increasing counterselection incubation time to 30 minutes in rounds 8 to 10, and by progressively increasing the amount of bead used. DNA libraries were analyzed throughout the selection process; first, rounds five to seven (CR5 to CR7) were sequenced, then, libraries from CR8 to CR10 were sequenced.

Aptamer selection for DnaK-NTD 2 (FN40R-10 Library)

For the second selection attempt for DnaK-NTD a different DNA library with 40-nucleotide random region was used (5' – AGA GCA TAA AGA CAG CCT CTAG – N40 –GAC CAA GAC GAA GGA CTC CATA – 3'). The forward primer used in both PCR 1 and PCR 2 was 5' –TAT GGA GTC CTT CGT CTT GGTC – 3', for PCR 1 the reverse primer was 5' –GAC CAA GAC GAA GGA CTC CATA – 3', and for PCR 2 the reverse primer was 5' – TTTTTTTTTTTTTTT /iSp18/ TAT GGA GTC CTT CGT CTT GGTC – 3'. DNA sequences were ordered from IDT, purified individually (thick dPAGE, 10 %), resuspended, and stored at -20 °C independently (library and primers). Surfaces were cleaned with 10 % bleach before and after each selection round, and pipetting was done with filter tips only. Tubes containing template were carefully handled and stored separately from primers and other reagents. Selection was started with 1000 pmol of library. Selection was conducted for eight rounds with 130 nM of library and target in selection buffer (SB, HEPES 50 mM, NaCl 150 mM, KCl, 6 mM, $MgCl_2$ 2.5 mM, CaCl 2.5 mM 0.01 % Tween-20). Counter selection against beads and BSA was introduced at round five. Stringency of selection was enhanced by increasing washing of beads from three washes and rinses progressively to eight. DNA pools after two rounds, and four to eight rounds of selection were analysed with deep sequencing. The libraries from this selection were termed NR1 to NR8.

2.5. Reselection of aptamers for full-length DnaK

Reselection was conducted for three rounds using the DnaK as the target protein. Reselection was conducted with the libraries for DnaK-CTD FN40R-7 (C-library) and for DnaK-NTD FN40R-10 (N-library). Reselection with the C-library started with the CR7 library, the following rounds were termed RS-C8, RS-C9 and RS-C10. Reselection with the N- library started with the NR5 library, the following rounds were termed RS-N6, RS-N7 and RS-N8. Reselection was conducted for three rounds, the selection reactions contained 100 pmol of library and bead-immobilized DnaK; 130 nM, 40 nM, and 20 nM in each progressive round (SB, HEPES pH 7 50 mM, NaCl 100 mM, KCl 10 mM, $MgCl_2$ 5 mM, 0.03 % Tween-20). The last two rounds of reselection were sent for deep-sequencing (RS-C9 and RS-C10, and RS-N7 and RS-N8).

2.6. Deep sequencing of DNA pools

Samples were prepared for sequencing on Illumina platforms to identify the composition of ssDNA pools after *in-vitro* selection. Illumina sequencing tags compatible with the primers of the libraries used were ordered and purified. For each round of selection, eight PCR reactions (50 µL) were prepared using PCR 1 (diluted 1:20) as the template, and the sequencing tags as primers. The PCR2 test was used as a reference for the number of amplification cycles. Samples were run on a 3 % agarose gel at 100 volts for ~ 1h. Bands were excised, and DNA was

purified using a GeneJET Gel Extraction Kit (Thermo Scientific). Samples were quantified using Nanovue and submitted for sequencing.

Illumina sequencing data was processed following the protocol set up by the lab manual. Primer sequences were removed, and the paired end reads were trimmed, merged, and clustered. The following software was used: Cutadapt 1.8.3 (Martin, 2011), usearch10.0.240_i86linux32 (Robert C. Edgar, 2010), FASTAtoTAB.py 1.2 (Jim Gu, Li Lab), ExtractTopN.py 1.0 (Jim Gu, Li Lab), IPLLv18.bash (Jim Gu, Li Lab). This analysis identified the identity of all sequences in the pool and their abundance. As such, sequences were ranked and top sequences and top families of sequences (clusters) were identified. Comparisons between rounds with the same library were done using Multi-Round Trend Processing and Multiple Sequence Alignment and Logo Generation (R. C. Edgar, 2004). Aptamer candidates were further analysed using the m-fold web server.

2.7. Characterization of aptamer binding

Labeling DNA with radioactive isotope

To visualize the behaviour of the aptamers, the DNA was radiolabeled at the 5' end. Aptamers and ssDNA pools (2 pmol) were incubated with ATP (γ^{32}P) and T4 PNK (T4 Polynucleotide Kinase) in 1x PNK buffer A at 37 °C for 1 hr (Thermo Scientific). Reaction was stopped by adding 2x urea loading buffer. Samples and a marker were incubated at 90 °C for 7 min and loaded onto a thin dPAGE (10%). The gel was run at 33 W for 20 min. Gel was wrapped and viewed

under UV light to identify the position of the marker. The corresponding band in the radioactive sample was excised and extracted using the gel crush and soak method (1h incubation at 60 °C in shaker, twice). DNA was recovered with ethanol precipitation and resuspended in SB such that 1 µL of sample contained at least 200 CPM. The estimated concentrations of radiolabeled DNA stocks were 10 to 3 nM, assuming a generous recovery rate of DNA (50 %). Radioactivity was tracked using Geiger counter, and all safety protocols were followed.

Electromobility shift assay (EMSA)

Binding reactions were set up to contain various known concentrations of target protein in SB. DNA library or aptamer was folded (90 °C for 5 min, ice for 10 min, room temperature for 15 min) and 2 µL were added to the reaction. The final volume of the reactions was 20 to 30 µL; thus, the concentration of DNA in sample was between 1 and 0.2 nM. Reaction was incubated at room temperature for one hour on a shaker. Then, 6X loading dye (bromophenol blue 1 g/L, xylene cyanole FF 1 g/L, 50 % glycerol) was added to the samples. Samples were loaded onto 7% native PAGE (mini), and run at 150 volts for 30 min. Gels were wrapped, and overall radioactivity was measured using a Geiger counter. Gels were exposed to storage phosphor plates for 5 to 45 min. Plates were imaged with Typhoon. Binding of aptamers to target was quantified using image-J by measuring the intensity of the shifted band of DNA and comparing it to the control.

Bead-based binding assays: Gels and Dot-blots

Protein targets were immobilized onto Ni-NTA magnetic beads as explained above and diluted to various known concentrations. Radiolabeled DNA was denatured and refolded in SB (90 °C for 5 min, ice for 10 min, room temperature for 15 min) and 2 µL were added to the reaction. Final volume of the reactions was 25 µL; thus, DNA concentration was estimated to be between 0.24 and 0.8 nM. Binding reactions were incubated for 1 h at room temperature while shaking. Separation of target-bound aptamers and unbound aptamers was done using a magnetic rack or a dot-blot assay.

Small-scale experiments used a magnetic rack to pellet the magnetic beads loaded with protein and target-bound aptamers **(Figure 2.2 A)**. The supernatant was removed and saved as the supernatant fractions. The beads were washed twice with SB. Bead fractions were resuspended in 25 µL of elution buffer (SB with 300 mM imidazole). Supernatant fractions (unbound aptamers) and bead fractions (bound aptamers) were run on a 10% dPAGE for 20 min at 33 Watts. Gels were wrapped and radioactivity was measured with a Geiger. Gels were exposed to phosphor storage plates for the appropriate length of time. Plates were imaged with Typhoon, and binding of aptamers was quantified with ImageJ.

A dot-blot assay uses membranes to separate the bound and unbound fractions of DNA **(Figure 2.2 B)**. The sample is spotted on a nitrocellulose (NC) membrane, bead-immobilized proteins, and any bound DNA bind to the NC layer while

unbound DNA flows through. Flow through is captured by a nylon membrane. The Minifold I system from Whatman was used for this experiment. The apparatus was assembled by connecting the vacuum pump to the vacuum plenum, then the filter support plate was placed on top. Membranes were soaked in SB and placed in the following order: two layers of blotting paper, nylon, and nitrocellulose at the top. Then, the sample well plate with the metal plate were placed on the top. Finally, the system was clamped firmly. The vacuum pump is used to pull the buffer down into the reservoir and dry the membranes. To load the samples, the vacuum was turned on and the wells were dried. The first row of wells was washed with 200 µL of SB (0.01% T-20) per well and dried. The samples were then added one by one while the vacuum was on. Samples were washed with 400 µL SB (0.01% Tween-20) and dried. The process was repeated row by row. Note that the SB used should be between 0.01% to 0.02 % Tween-20, decreasing the content in the buffer may lead to nonspecific binding of aptamers to the NC and increasing it leads to bubbling of the buffer in the manifold. Once all samples were loaded and dried, the apparatus was disassembled, and the NC and nylon membranes were placed between transparency film. Overall radioactivity was measured using a Geiger counter, and membranes were exposed to storage phosphor plates. Plates were imaged with Typhoon. Binding was quantified by measuring the intensity of the radioactive DNA shadow on the NC membrane and comparing it to the total intensity on both the NC and nylon membranes.

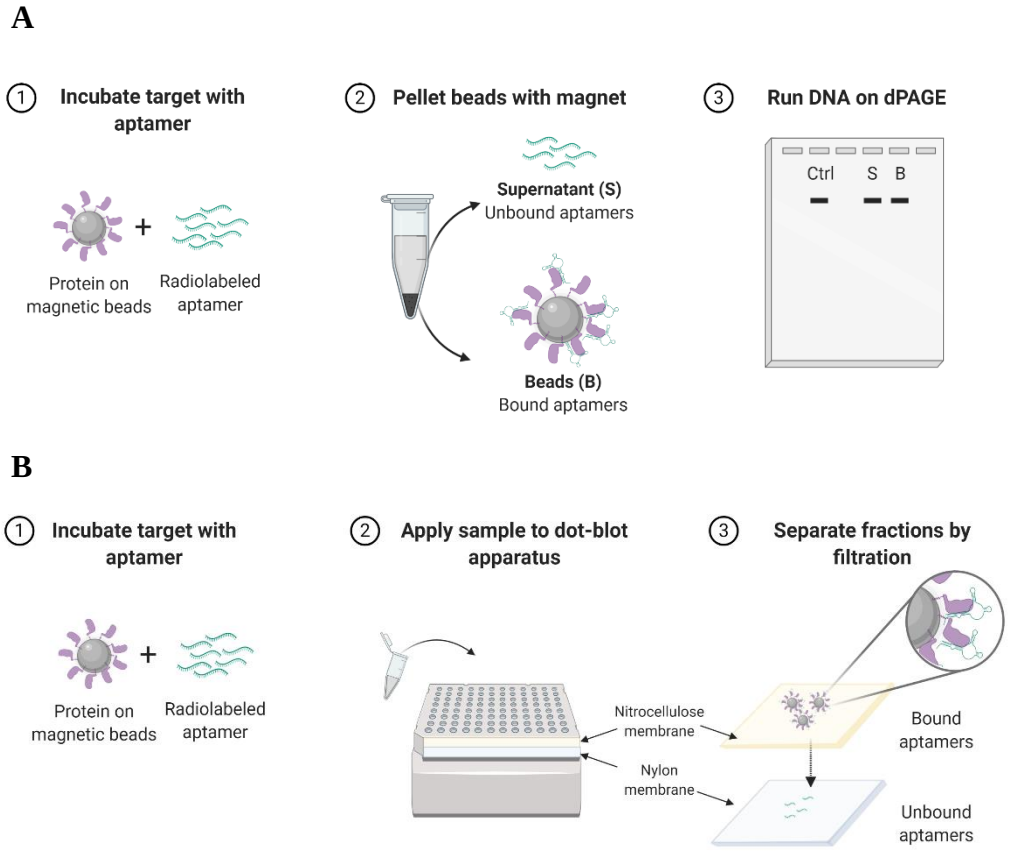


Figure 2.2 Testing aptamer binding with bead-based assays. Aptamer binding to the target was tested with bead-based assays. Radiolabeled aptamers were incubated with bead-immobilized protein for 1h at RT. **(A)** In small scale experiments beads were pelleted with a magnet to separate the supernatant (S) with unbound aptamers, and the beads (B) with protein and bound aptamers. Fractions were run on 10% dPAGE. Gels were exposed to storage phosphor plates and imaged. **(B)** For large scale experiments a dot-blot apparatus with 96 sample wells was used. Sample was applied on nitrocellulose (NC) membrane and washed with buffer. Beads with bound protein and aptamers stayed on the NC membrane, while the unbound aptamer flowed through to the nylon membrane. A vacuum pump was used to drain the buffer and dry the membranes. Membranes were exposed to storage phosphor plates and imaged.

CHAPTER 3:PURIFICATION OF TARGET PROTEINS

3.1. Designing expression vectors for *Fusobacterium nucleatum* genes *dnak, dnak-ctd and dnak-ntd.*

While the structure of *F. nucleatum* DnaK has not been resolved, its structure has been inferred by homology using the Swiss-model server. All DnaK proteins share a similar modular structure; two domains connected by a highly conserved 18-amino acid linker. The protein was divided at the linker region leaving each domain with nine linker amino acids. The gene for *F. nucleatum dnak* was synthesized by Biomatik and introduced into a pET15b expression vector, which adds an N-terminal His-tag to the proteins. The plasmids for *dnak, dnak-ctd* and *dnak-ntd* were introduced into two *E. coli* strains: BL21 cells were used for protein expression, and the DH5α cells were used to store and proliferate the plasmid. Overall, the high purity of the plasmid mix made all the transformation attempts successful as colonies always formed; however, the colony number varied from a few colonies (three to ten) to dozens.

To verify the gene sequences in the plasmids, *dnak, dnak-ctd* and *dnak-ntd* plasmids were extracted from DH5α cells and sequenced from the T7 promoter and the T7 terminator. Sequencing data for the plasmids showed 100% DNA sequence identity to the codon optimized gene given by Biomatik. Thus, these plasmids were verified to have complete amino acid sequence identity to the *F. nucleatum* DnaK protein.

3.2. Expression and purification of target proteins DnaK, DnaK-NTD and DnaK-CTD

Proteins were expressed using *E. coli* BL21 cells compatible with the pET15b expression vector. For large scale expression of the proteins of interest, cells were cultured in 1 L of LB-Amp and grown to an OD_{600} of 0.7, and expression was induced with IPTG for four hours at 37 °C. Expression of DnaK, DnaK-NTD and DnaK-CTD did not appear to produce cytotoxicity because induced cells had similar grow rates to uninduced cells.

Cell lysate was first purified through a HisTrap HP Ni-NTA affinity column, which binds histidine residues on proteins. As such, the His-tagged target proteins bind tightly to the column. Proteins are eluted from the His-column by washing with buffer containing imidazole; thus, weakly bound proteins elute off the column with low concentration of imidazole while tightly bound proteins elute at higher concentrations of imidazole. Selected fractions containing the protein of interest were purified through HiTrap Q HP anion exchange column, which binds negatively charged proteins. As a result, proteins are eluted according to their isoelectric point (PI) using a gradient of increasing concentrations of NaCl. To ensure the proteins of interest are negatively charged, the buffers used had a pH of 8, much higher than the PI of recombinant proteins DnaK (5.18), DnaK-NTD (5.82) and DnaK-CTD (5.15).

Eluted fractions with high protein concentrations are detected by the UV 280 nm sensor on the machine, and their protein content were checked by SDS-PAGE (12%). DnaK purified by the HisTrap was eluted mostly between 90 mM to 100 mM of imidazole **(Figure 3.1 A)** (F12-F16), and it was eluted from the anion exchange column at between 290 to 340 mM of NaCl **(Figure 3.1 B)** (F13-F15). DnaK-NTD was eluted from the HisTrap between 80 mM to 85 mM imidazole **(Figure 3.2 A and B)** (F17-F21), and elution from the anion exchange column peaked at 260 mM NaCl **(Figure 3.2 C)** (F6-F9). On the other hand, DnaK-CTD eluted from the HisTrap over a wide range of imidazole (50 mM to 250 mM) **(Figure 3.3 A)** (F5-F35); thus, the fractions needed to be concentrated before purification through the Q column. DnaK-CTD showed peak elution between 280 and 350 mM of NaCl from the anion exchange column **(Figure 3.3 B and C)** (F10-F12). The three proteins of interest were successfully isolated and concentrated into stocks. The purity of the proteins was tested by separation of the protein stocks in an SDS-PAGE. The identity of the proteins was confirmed by size, and the presence of a single band at the expected size verified the purity of the protein **(Figure 3.4)**.

3.3. Discussion and conclusion

The multidomain selection approach requires the target protein to be split into its domains and to conduct selection on each domain individually. DnaK (Hsp70) is a modular protein with two functional domains connected by a linker peptide made up of mostly hydrophobic residues. The N-terminal domain known as the nucleotide-binding domain (NBD) has an ATP binding pocket and ATPase activity (Flaherty et al., 1990). The C-terminal domain, known as the substrate-binding domain (SBD), recognizes and binds unfolded polypeptides (Zhu et al., 1996). Structural studies have demonstrated that DnaK has a conformational cycle; DnaK opens when bound to ATP and has low affinity for peptide substrates; then, upon ATP hydrolysis, DnaK closes and has high affinity to bound peptides (Kityk et al., 2012; M. P. Mayer & Bukau, 2005; Matthias P. Mayer, 2013). Studies have also demonstrated that when DnaK is free or bound to ADP the domains behave independently of each other; thus, the domains can be separated and still maintain functionality (Bertelsen et al., 2009; Buchberger et al., 1995). As a result, DnaK appears to be an appropriate protein target for testing the multidomain selection method.

Separate expression vectors were designed for DnaK, DnaK-NTD and DnaK-CTD. The *dnak-ntd* vector contained the NBD and nine residues of the linker region, while *dnak-ctd* encoded the SBD and the remaining nine residues of the linker. All proteins were N-terminally His-tagged. The *E. coli* BL21 expression system consistently produced the protein targets effectively (~ 200 µg of protein

per litre of culture). Purification of the full-length DnaK required careful monitoring and selection of fractions to optimize for the presence of DnaK and not other proteins; DnaK has the SBD in the CTD that readily associates with other proteins. Hence, once these contaminant bands are present in solution, they are very difficult to separate. On the other hand, DnaK-NTD does not have a peptide binding pocket and it was easily purified in high quantities. Purification of DnaK-CTD was the most challenging because the His-tag was not properly displayed on the protein; thus, it eluted from the His-column over a wide range of imidazole concentrations. Then, 30 fractions were pooled for separation with the anion exchange column. To avoid this issue, DnaK-CTD should be His-tagged at the C-terminus in future designs. The isolated protein stocks were scrutinized by SDS PAGE analysis and only highly pure stocks were chosen for selection. Overall, this method yielded highly pure protein stocks that were subsequently used for aptamer selection.

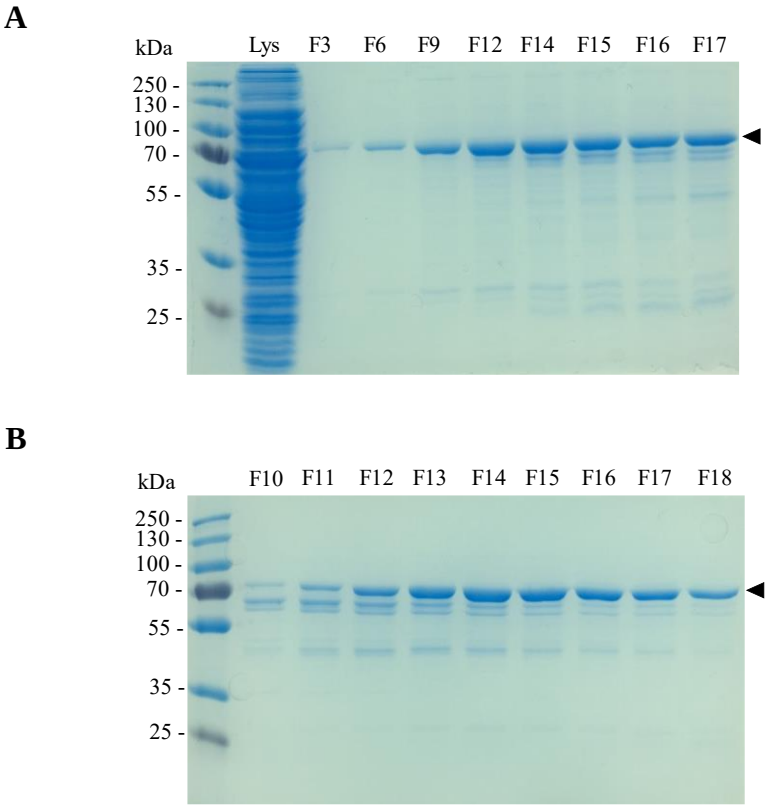


Figure 3.1 Purification of *F. nucleatum* DnaK. DnaK (67.6 kDa) was expressed on E. coli BL21 cells. **(A)** Cell lysate (Lys) was purified through a HisTrap HP Ni-NTA affinity column. Lysate and samples from fractions F3 to F17 were run on a 12 % SDS PAGE. Fractions F12 to F16 were pooled. **(B)** Protein was purified through a HiTrap Q HP anion exchange column. Samples from fractions 10 to 18 were run on a 12% SDS PAGE. SDS gels were stained with Coomassie blue. Black arrows (◄) indicate DnaK bands. F represents the elution fraction number of each sample, 20 µL of each fraction were used; increasing fraction number corresponds to higher concentration of elution buffer.

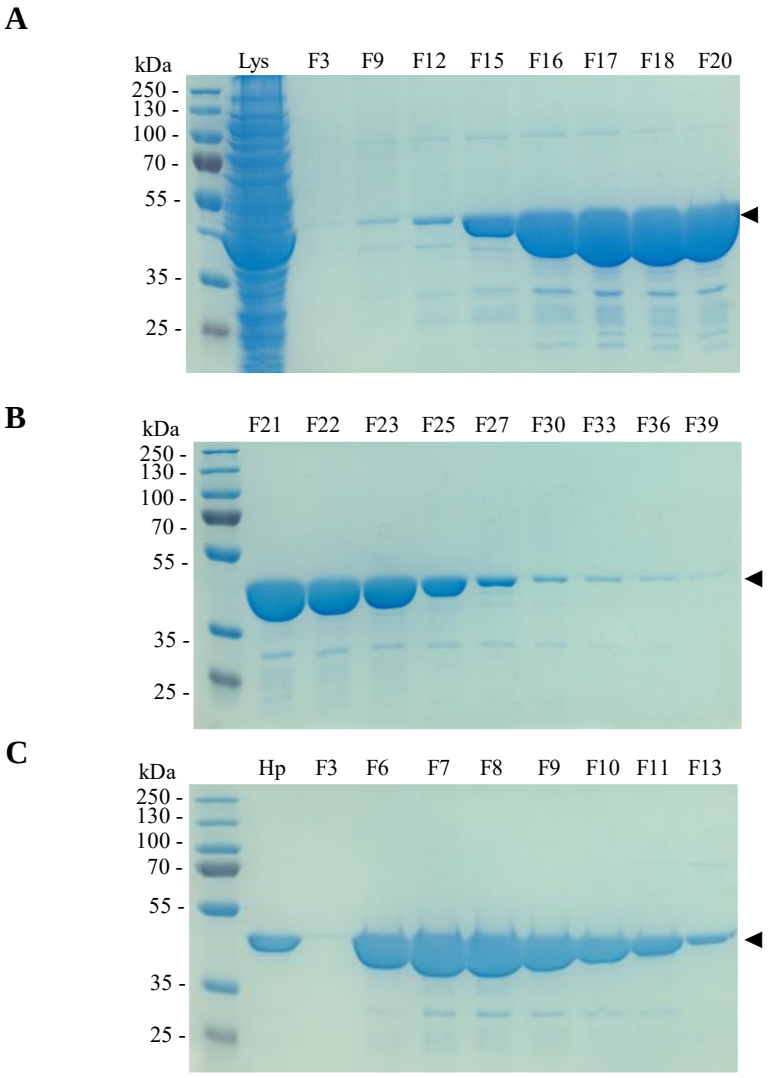


Figure 3.2 Purification of *F. nucleatum* DnaK-NTD. DnaK-NTD (40.5 kDa) was expressed on E. coli BL21 cells. Cell lysate (Lys) was purified through a HisTrap HP Ni-NTA affinity column. **(A)** Lysate and samples from fractions F3 to F20, and **(B)** F21 to F39 were run on a 12 % SDS PAGE. **(C)** Protein from F17to F21 were pooled (Hp) and purified through a HiTrap Q HP anion exchange column. Samples from fractions F3 to F13 were run on a 12% SDS PAGE. SDS gels were stained with Coomassie blue. Black arrows (◀) indicate DnaK-NTD bands. F represents the elution fraction number of each sample, 20 µL of each fraction were used; increasing fraction number corresponds to higher concentration of elution buffer.

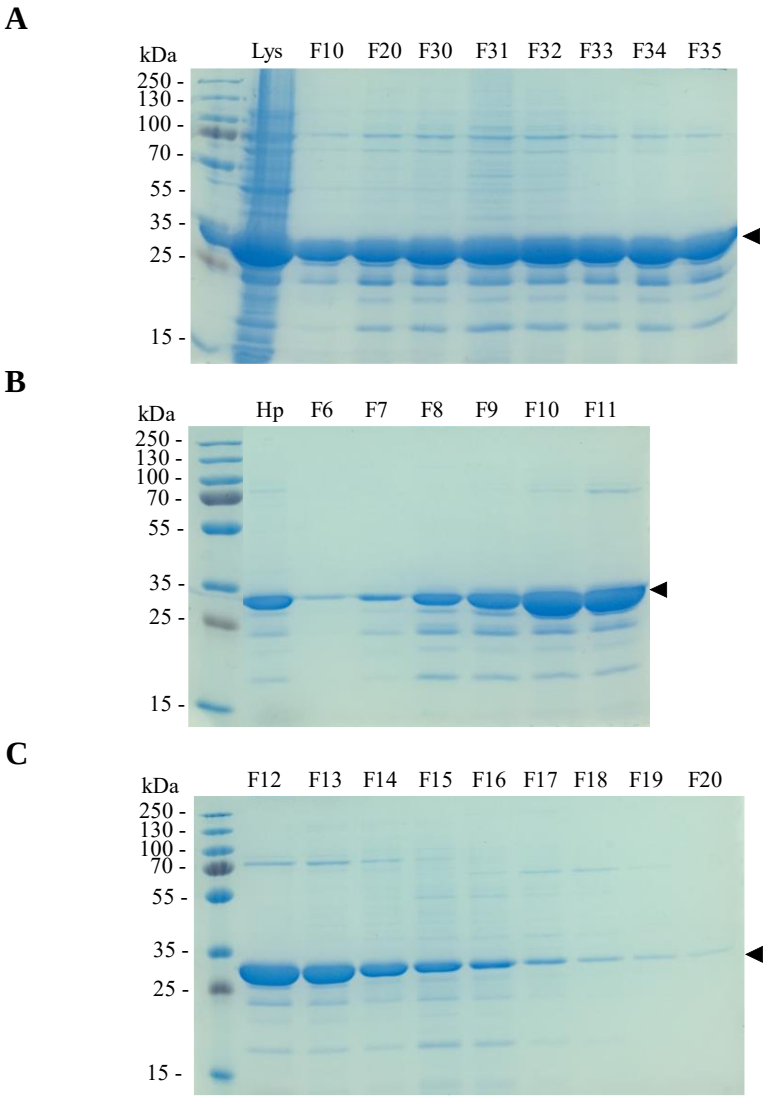


Figure 3.3 Purification of *F. nucleatum* DnaK-CTD. DnaK-CTD (29.3 kDa) was expressed on *E. coli* BL21 cells. Cell lysate (Lys) was purified through a HisTrap HP Ni-NTA affinity column. **(A)** Lysate and samples from fractions F10 to F35 were run on a 12 % SDS PAGE. **(B)** Protein from F10 to 30 were pooled (Hp) and purified through a HiTrap Q HP anion exchange column. Samples from fractions F6 to F11, and **(C)** F12 to F20 were run on a 12% SDS PAGE. SDS gels were stained with Coomassie blue. Black arrows (◀) indicate DnaK-CTD bands. F represents the elution fraction number of each sample, 20 µL of each fraction were used; increasing fraction number corresponds to higher concentration of elution buffer.

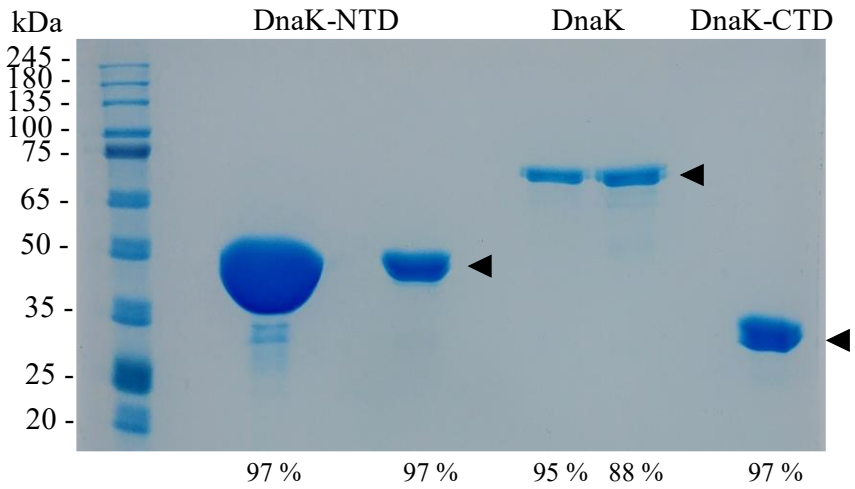


Figure 3.4 Purified *F. nucleatum* DnaK, DnaK-NTD and DnaK-CTD. Proteins were purified using a HisTrap HP Ni-NTA affinity column, followed by an anion exchange column (HiTrap Q HP). Proteins were concentrated with centrifugal filters and 1 µL of concentrate was run on a 12 % SDS PAGE. Gel was stained with Coomassie blue. Black arrows (◀) indicate the band of the target proteins. Purity of each protein was calculated with ImageJ and it is presented below each lane. Each lane represents a different purification experiment.

CHAPTER 4:APTAMER SELECTION FOR *F. NUCLEATUM* DNAK N-TERMINAL DOMAIN (DNAK-NTD)

4.1. Selection of aptamers for DnaK-NTD

Aptamers for DnaK-NTD were selected using ten rounds bead-based SELEX **(Figure 2.1)**. Libraries from rounds seven to ten (N7 to N10) were used for deep sequencing analysis, which provides the identity of the sequences in the DNA pool and their relative abundance. Similar sequences were grouped into clusters where members have up to four different nucleotides. A summary of the sequencing data for rounds N7 to N10 is presented in **(Table 4.1)**. These data showed strong evidence of sequence enrichment because the number of unique clusters in the pool decreased, and the average frequency of clusters increased with each round of selection **(Table 4.1)**.

To identify aptamer candidates, the top 10 clusters after ten rounds of selection were examined **(Table 4.2)**. At N10, 88.19 % of the DNA pool was composed of the top four clusters. Clusters 5 to 8 were present in considerably lower abundance and made up 0.29 % of the library. The remaining clusters were sparsely represented in the DNA pool and had a low number of reads. Therefore, these data suggest that the DNA pool is significantly enriched towards few selected sequences at N10. Analysis of enrichment pattern for the top four clusters was conducted by studying the sequencing data from N7, N8, N9 and N10 **(Figure 4.1)**. The top four clusters were found to progressively increase in relative abundance between round N8 and N10 **(Figure 4.1).** Overall, the

enrichment patterns suggest that these clusters may contain suitable aptamer candidates. Thus, four aptamers corresponding to the top four clusters were identified as candidates and called NAp 1, NAp 2, NAp 3 and NAp 4.

Table 4.1 Summary of sequencing data for N7 to N10. Sequencing data from DNA pools after 7, 8, 9 and 10 rounds of selection against *F. nucleatum* DnaK-NTD. The number of sequences read by the equipment for each pool is represented as the total number of reads. The number of unique clusters per round was calculated by organizing the individual reads into unique clusters (families of similar sequences). Calculated average number of reads per cluster and the average frequency of clusters, and the frequency of the most abundant cluster in each round is presented.

Round	Total Number of Reads	Number of Unique Clusters	Average Number of Reads per Cluster	Average Frequency of Clusters (%)	Frequency of Top Cluster (%)
N10	298705	34225	8.728	2.92×10^{-3}	67.3
N9	268845	169692	1.584	5.89×10^{-4}	31.6
N8	200537	188033	1.066	5.32×10^{-4}	5.2
N7	251900	230986	1.091	4.33×10^{-4}	7.4

Table 4.2 Top 10 clusters at N10. After ten rounds of selection against *F. nucleatum* DnaK-NTD the DNA pool has become enriched. The top 10 families of sequences are presented, and their relative abundance is described as a percentage.

Cluster	Abundance	Sequence
1	67.27	CGCATGGTCTTGTACTCGTGCGGGCCGTCTGAGGGAAGTT
2	12.08	CGGAGCAGGTCCAGGCCATTTAATTGTGCTTACCGTAGAG
3	4.47	GAGGCCGAGAAGTGCCGAGGCTTGTTTAACAAGGGGCTGC
4	4.38	GTGTGCGAGAAGACACACGGCATGTCTGAATGGTCCGTTG
5	0.10	CGCGAAAGACGTGCGCCGCGGGGCCCGCGCGGCCGTTCGG
6	0.09	TGGTAGTGCATGCCATCTACCAGGCGCGTCTGAGTGTGGC
7	0.06	CGCTTTTGACGTGCCCGCTCCGTCAGGTAGAGGCTCAGGG
8	0.04	CTTTGATGAGGTCGGCGAAAGACCGACGGCGCTAACGTCA
9	0.01	CGCGTGATGACTATAAGACGTGGCCGAAGAGCCCCAGGG
10	0.01	GTGTGCGAGAAGTCACACGGCATGTCTGAATGATTTGTTG

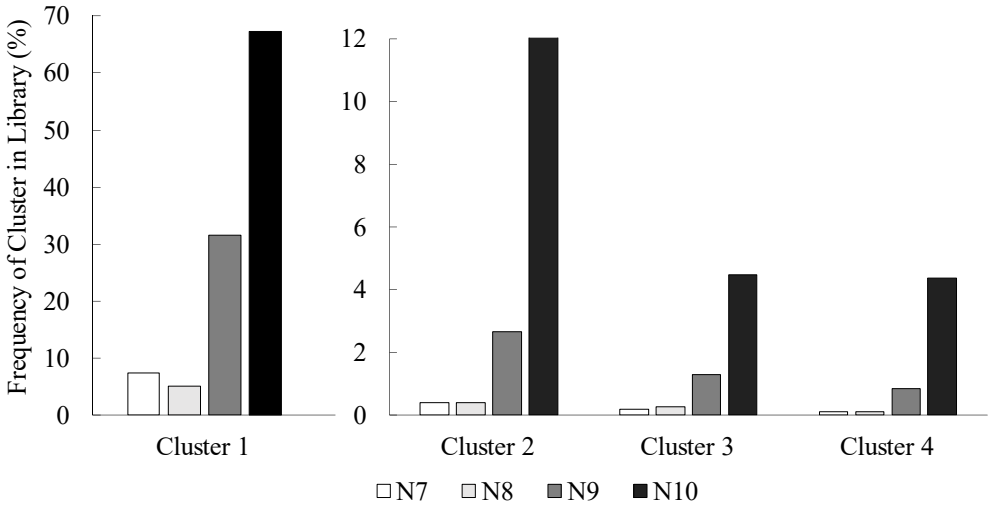


Figure 4.1 Abundance of top four sequence clusters in N7, N8, N9 and N10. The relative abundance of the top four clusters was tracked from round seven to ten and presented as frequency in the DNA pool as a percentage. Clusters increase in abundance over each round of selection, especially after round nine.

4.2. Characterization of aptamer candidates

Aptamer candidates NAp 1, NAp 2, NAp 3 and NAp 4 were characterized using electromobility shift assays (EMSAs). Each aptamer was incubated with different concentrations of DnaK-NTD (0 to 5 μM) for one hour. Then, the protein-aptamer mixtures were separated using a native PAGE. Binding of the aptamers to the target was detected as a shift of the DNA band upwards on the gel; the protein-aptamer complex moves slower than the aptamer alone. Binding was quantified using Image J. Only aptamer NAp 1 aptamer bound significantly to DnaK-NTD **(Figure 4.2)**. Further characterization of NAp 1 demonstrated consistent binding

to DnaK-NTD **(Supplemental Figure 1A)**. Binding of NAp 1 to DnaK-NTD was quantified and a binding curve was plotted; this curve suggests that 50 % of the aptamer binds the target between 3 and 4 µM **(Supplemental Figure 1 B)**. The data was analysed with GraphPad Prism using non-linear regression for one-site binding ligand. The dissociation constant (Kd) of NAp 1 was estimated to be 13.23 µM (R^2 0.9254).

Binding of NAp 1 was also tested using a bead-based assay. Here, DnaK-NTD was loaded onto beads and diluted to various concentrations (0 to 10 µM) and incubated with NAp 1. The beads were pelleted with a magnetic rack, and the supernatant with unbound aptamers was removed. The pelleted protein-aptamer complex was resuspended in buffer. All fractions were run on a denaturing PAGE. Gels were exposed to phosphor plates for imaging. NAp 1 was found to weakly bind bead-immobilized DnaK-NTD **(Figure 4.3)**. Quantification of binding was done with ImageJ and plotted over concentration of DnaK-NTD. Binding data (N=1) was analysed with GraphPad Prism and fitted to a non-linear fit for a one-site binding ligand. The Kd of NAp1 with the bead-binding assay was estimated to be 3.78 µM (R^2 0.8884) (**Supplemental Figure 2)**.

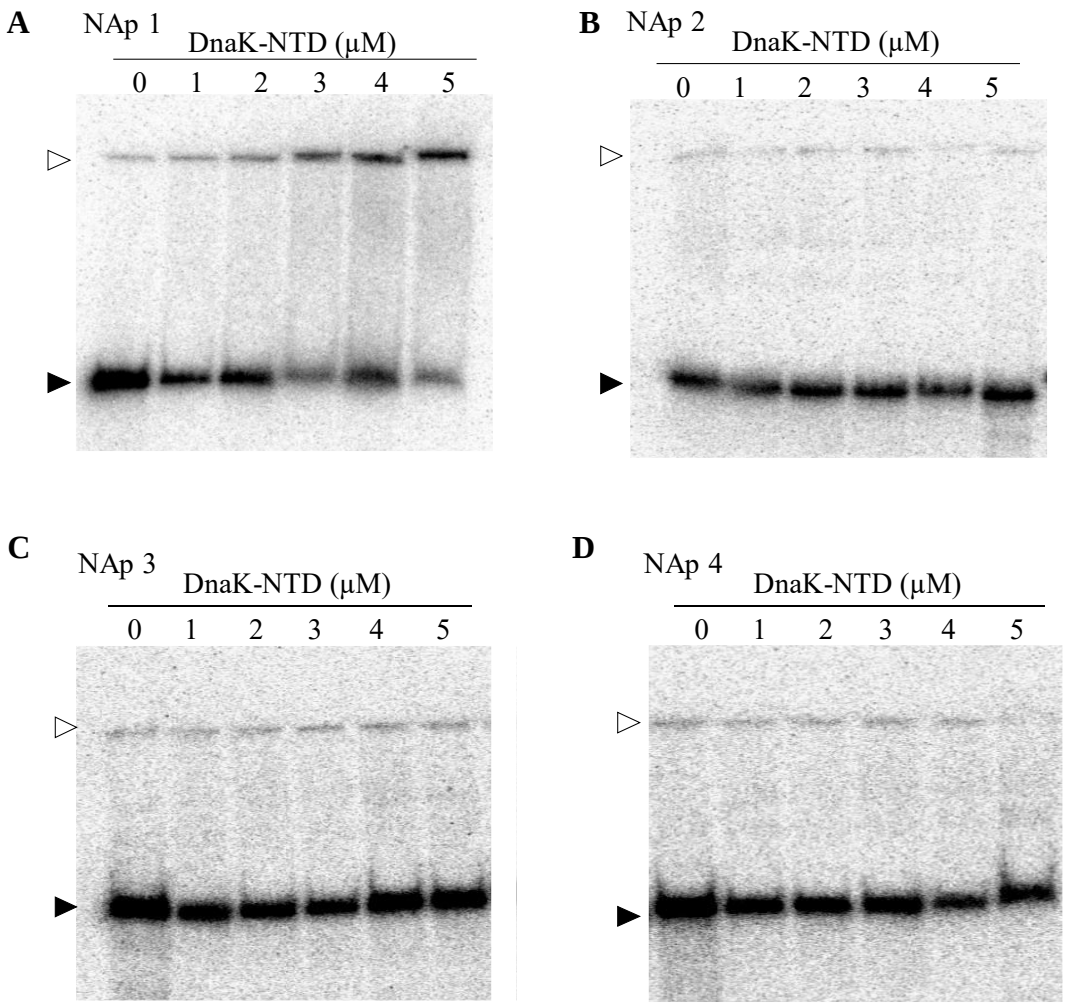
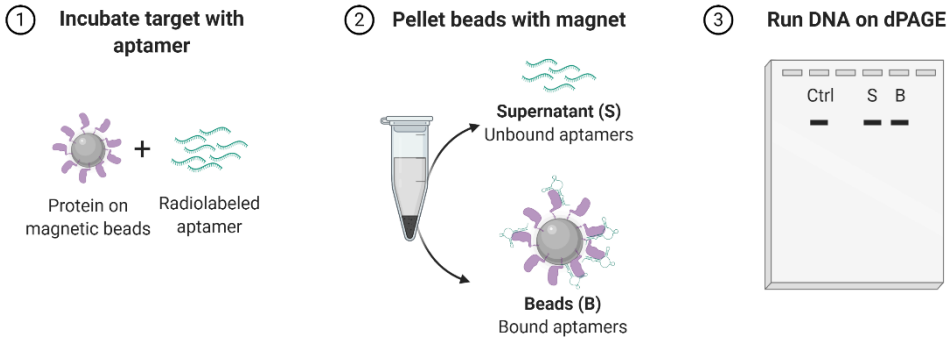


Figure 4.2 Binding test of aptamers NAp 1, NAp 2, NAp 3 and NAp 4 to F. nucleatum DnaK-NTD with electromobility shift assay (EMSA). EMSA was performed with 25 nM of aptamer **(A)** NAp 1, **(B)** NAp 2, **(C)** NAp 3, or **(D)** NAp 4, and various concentration of DnaK-NTD protein (0 μM to 5 μM). Aptamers were radiolabeled, incubated with target for one hour, and run on native PAGE (7 %). Gels were exposed to phosphor plates for 7 minutes and scanned with Typhoon. Unbound aptamer is indicated by the black arrowhead (►), and the wells are indicated by the white arrowhead (▷). Binding of aptamer was calculated with ImageJ. NAp 1 bound to *F. nucleatum* DnaK-NTD as noted by the aptamer-protein complex in the wells.

A



B

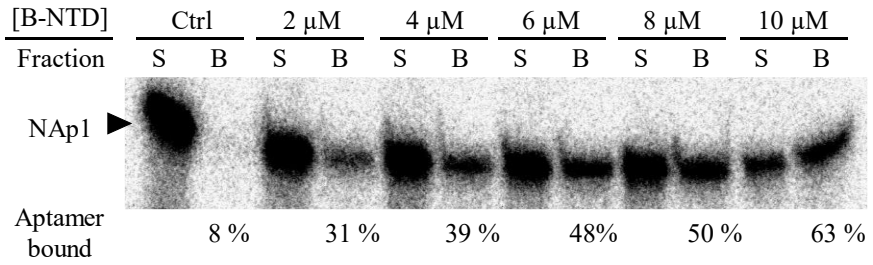


Figure 4.3 Binding test of NAp 1 against *F. nucleatum* DnaK-NTD. Binding of NAp 1 to DnaK-NTD was tested using a bead-binding test. DnaK-NTD was immobilized on magnetic beads (B-NTD) and diluted to various concentrations (2 to 10 µM). Radiolabeled aptamers were incubated with B-NTD for one hour. A magnetic rack was used to pellet the beads and the supernatant (S) with unbound aptamers was removed. Beads with bound aptamers (B) were resuspended in buffer. Supernatant and bead fractions were run on a denaturing urea PAGE (10 %). **(A)** Schematic of experimental design. **(B)** Gels were exposed to phosphor plates and imaged with Typhoon. Binding of aptamer was quantified using ImageJ.

4.3. Comparative analysis of DnaK-NTD selection: *F. nucleatum* vs. *C. difficile*

Previously in the Li Lab, aptamer selection for DnaK-NTD of *Clostridium difficile* was performed using the same FN40R-6 library. The DNA libraries of DnaK-NTD aptamer selection for *F. nucleatum* round 10 (F.n. N10) and *C. difficile* round 9 (C.d. N9) were compared. The top eight clusters at F.n. N10 were identified to have enriched through selection in section 4.2. However, they were also present in the C.d. N9 library and made up 91.4 % of all C.d. N9 sequences **(Figure 4.4)**. Clusters 1 to 4 were present with rather similar frequencies in both libraries, while clusters 5 to 8 were more prevalent in the C.d. N9 library **(Figure 4.4)**. The two libraries were substantially similar and F.n. N10 did not contain any unique classes of sequences. Moreover, the DNA pool of C.d. N9 is enriched more uniformly towards the selected sequences than the library of F.n. N10. Early rounds were analysed to further characterize the source of the DNA cross-contamination. At round 2, NAp 1 was significantly enriched at 10 copies **(Supplemental Figure 3 A)**. NAp 1, NAp 2, NAp 3 and NAp 4 were also found to be significantly enriched at round 4 and 6 **(Supplemental Figure 3 B and C)**.

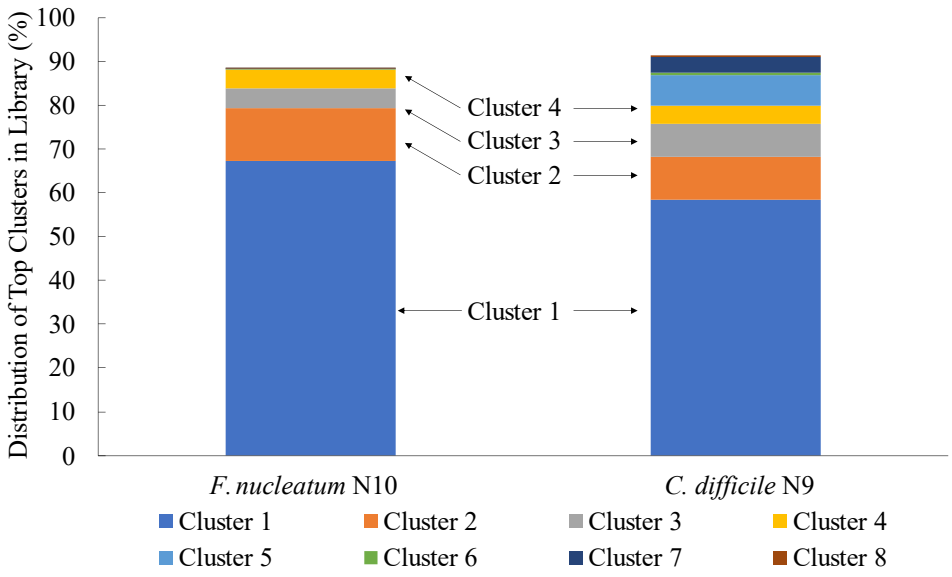


Figure 4.4 Distribution of top 8 clusters in N-libraries for *F. nucleatum* vs. *C. difficile*. Top 8 clusters identified after 10 rounds of selection against *F. nucleatum* DnaK-NTD (F.n. N10), were mapped to the *C. difficile* DnaK-NTD DNA library after 9 rounds (C.d. N9). The bar graph shows the frequency of the top 8 clusters as a percentage of the F.n. N10 and C.d. N9 libraries. The same top 8 clusters were found to make up 88.5 % of F.n. N10, and 91.4 % of C.d. N9.

4.4. Second selection of aptamers for DnaK-NTD

Selection of aptamers for DnaK-NTD was repeated using a different library to overcome the DNA cross-contamination problem. Selection was conducted for eight rounds following with a bead-based SELEX protocol and the FN40R-10 ssDNA library **(Figure 2.1)**. Multiple precautions were implemented to avoid DNA contamination. New stocks of library and primer sequences were ordered. Sequences were purified separately and stored in separate boxes. Tubes containing template were handled with utmost caution. To avoid cross-contamination, all surfaces and equipment were regularly cleaned with 10 %

bleach and the selection protocol was performed using filter tips. After eight rounds of selection, DNA pools were sent for deep sequencing to monitor the progression of selection.

Sequencing data was processed and organized by the relative abundance of individual sequences in each pool of DNA. Sequences were also grouped into clusters and ranked by abundance. The DNA pools showed substantial evidence of enrichment between rounds NR7 and NR8; the number of unique clusters drastically decreased and the average frequency of the clusters increased **(Table 4.3)**. Furthermore, the top sequence made up over 70 % of both DNA pools. Then, rounds NR2, NR4, NR5, and NR6 were also sent for deep sequencing analysis to conduct a more comprehensive examination of the enrichment pattern. Data from rounds two to six provided strong evidence of enrichment: unique clusters steadily declined and the average read per cluster increased over time **(Table 4.3).**

The top 10 clusters at NR7 were analysed to identify aptamer candidates **(Table 4.4)**. Cluster 1 represented 75.9 % of the library. Clusters 2 to 5 were also noticeably enriched and composed 22.7 % of the library. Cluster 6 and 7 were moderately enriched and made up 0.7 % of the library. All other sequences were present in low copy numbers and made up the remaining 0.7 % of the library. Therefore, the DNA pool at NR7 was enriched towards few selected sequences. Enrichment pattern of the last three rounds shows that the top four clusters enriched steadily from NR6 to NR8 **(Figure 4.5)**. As a result, the top four clusters were chosen as aptamer candidates and termed N7-1, N7-2, N7-3, and N7-4.

Table 4.3 Summary of sequencing data for NR2 to NR8. Sequencing data from DNA pools after two to eight rounds of selection against *F. nucleatum* DnaK-NTD. The number of sequences read by the equipment for each pool is represented as the total number of reads. The number of unique clusters per round was calculated by organizing the individual reads into unique clusters (families of similar sequences). Calculated average number of reads per cluster and the average frequency of clusters, and the frequency of the most abundant cluster in each round is presented.

Round	Total Number of Reads	Number of Unique Clusters	Average Number of Reads per Cluster	Average Frequency of Clusters (%)	Frequency of Top Cluster (%)
NR8	249927	117	2,136.13	0.88547	71.81
NR7	350588	2297	152.63	4.35×10^{-2}	75.89
NR6	814748	439732	1.85	4.53×10^{-4}	29.42
NR5	716800	709419	1.01	1.41×10^{-4}	0.43
NR4	955735	953826	1.00	1.05×10^{-4}	0.11
NR2	833261	832520	1.00	1.20×10^{-6}	2.78×10^{-3}

Table 4.4 Top 10 sequence clusters at NR7. After seven rounds of selection against *F. nucleatum* DnaK-NTD the DNA pool has become enriched. The top 10 families of sequences are presented, and their relative abundance is described as percentage.

Cluster	Abundance	Sequence
1	75.89	CGCATGGTCTTGTACTCGTGCGGGCCGTCTGAGGGAAGTT
2	9.46	CGGAGCAGGTCCAGGCCATTTAATTGTGCTTACCGTAGAG
3	6.33	GAGGCCGAGAAGTGCCGAGGCTTGTTTAACAAGGGGCTGC
4	5.06	GTGTGCGAGAAGACACACGGCATGTCTGAATGGTCCGTTG
5	1.82	CGCGAAAGACGTGCGCCGCGGGGCCCGCGCGGCCGTTCGG
6	0.55	TGGTAGTGCATGCCATCTACCAGGCGCGTCTGAGTGTGGC
7	0.14	CGCTTTTGACGTGCCCGCTCCGTCAGGTAGAGGCTCAGGG
8	0.01	CTTTGATGAGGTCGGCGAAAGACCGACGGCGCTAACGTCA
9	0.01	CGCGTGATGACTATAAGACGTGGCCGAAGAGCCCCAGGG
10	0.01	GTGTGCGAGAAGTCACACGGCATGTCTGAATGATTTGTTG

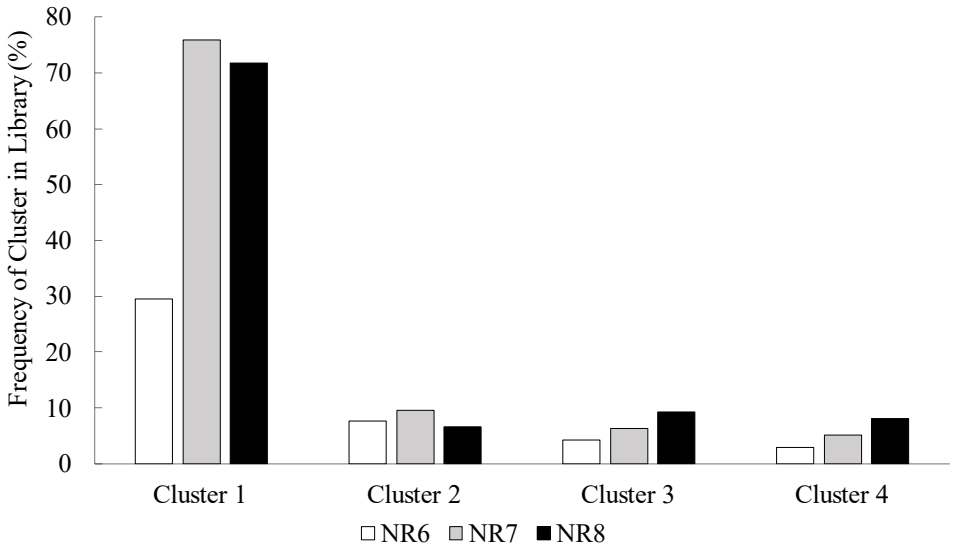


Figure 4.5 Abundance of top four clusters at NR6, NR7 and NR8. The top four clusters at NR7 were identified and their frequencies were tracked to previous rounds of selection. The frequency of the clusters is presented as a percentage of their abundance in the pool. Clusters 1 to 4 significantly increase in abundance after round six of selection.

4.5. Assessment of binding of aptamer candidates against DnaK-NTD

Binding of aptamers N7-1, N7-2, N7-3, and N7-4 to DnaK-NTD was assessed using electromobility shift assays (EMSAs). Each aptamer was incubated with different concentrations of DnaK-NTD (0 to 2 µM), and then the mixture was separated using a native PAGE electrophoresis. The N7 aptamers did not bind to DnaK-NTD as they did not produce a shift of the DNA band **(Figure 4.6)**.

The aptamers were tested under the same conditions of selection to assess whether aptamer binding is dependent on the protein's conformation while immobilized on the magnetic beads. Thus, aptamers were incubated with DnaK-NTD loaded on beads (B-NTD) in a one-to-one ratio (130 nM). Aptamers were also incubated with bead-immobilized DnaK-CTD (B-CTD) and with free beads (B) to control against non-specific binding. After incubation, beads were aggregated using a magnetic rack to separate the unbound fractions (supernatant) and bound fractions (beads with associated protein and aptamers) **(Figure 4.7)**. The supernatant was removed, and the beads were washed. Both the unbound fractions and bound fractions were run on a dPAGE. Aptamers N7-1, N7-2, N7-3, and N7-4 were found to bind bead-immobilized B-NTD specifically, while they did not bind to B-CTD **(Figure 4.7)**. Minimal background binding to the free beads was detected **(Figure 4.7)**. Quantification of aptamer binding using Image J showed that 40 to 48 % of the aptamers were bound to the target B-NTD.

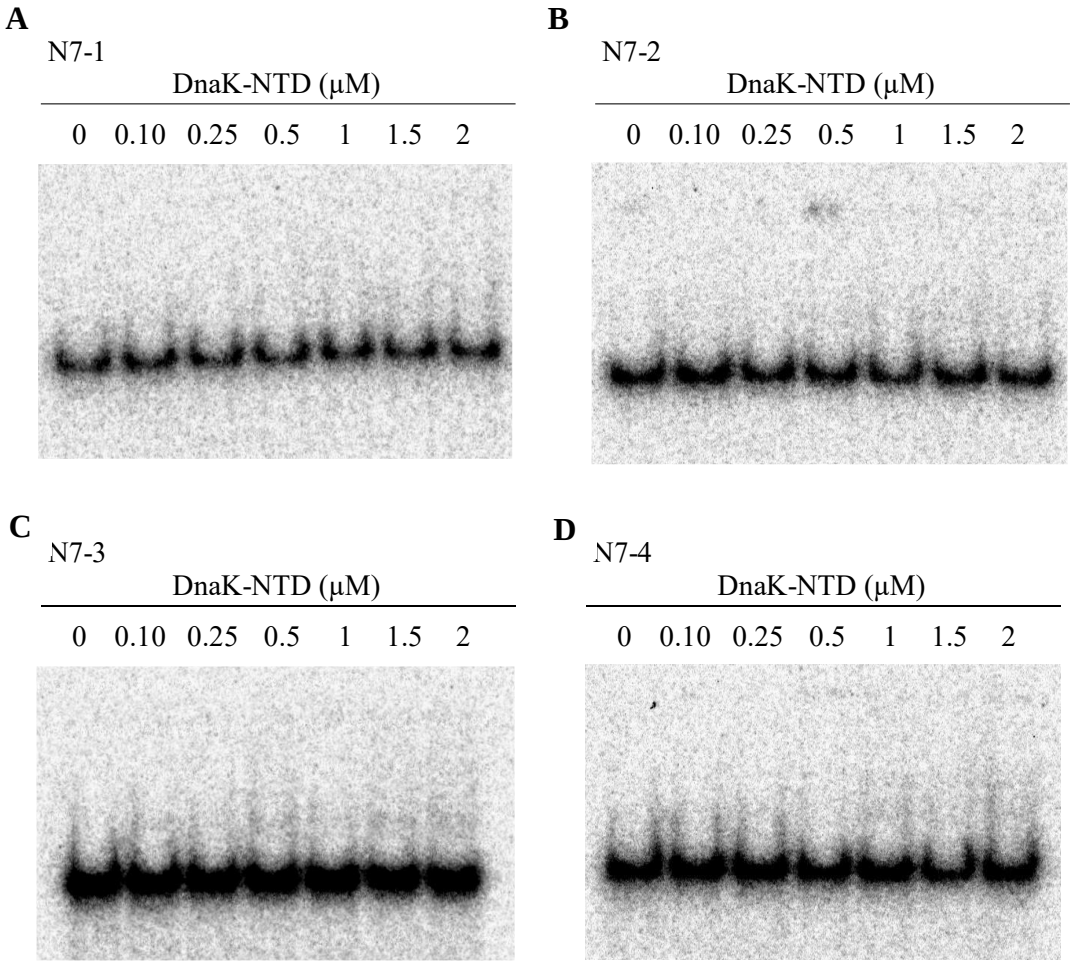


Figure 4.6 Binding test of N7 aptamers to *F. nucleatum* DnaK-NTD with electromobility shift assay (EMSA). EMSA was performed with aptamers **(A)** N7-1, **(B)** N7-2, **(C)** N7-3, or **(D)** N7-4, and various concentration of DnaK-NTD protein (0 μM to 2 μM). Aptamers were radiolabeled, incubated with target for one hour, and run on a native PAGE (7 %). Gels were exposed to phosphor plates and imaged with Typhoon. Binding of aptamers was calculated with ImageJ. N7 aptamers did not bind to *F. nucleatum* DnaK-NTD.

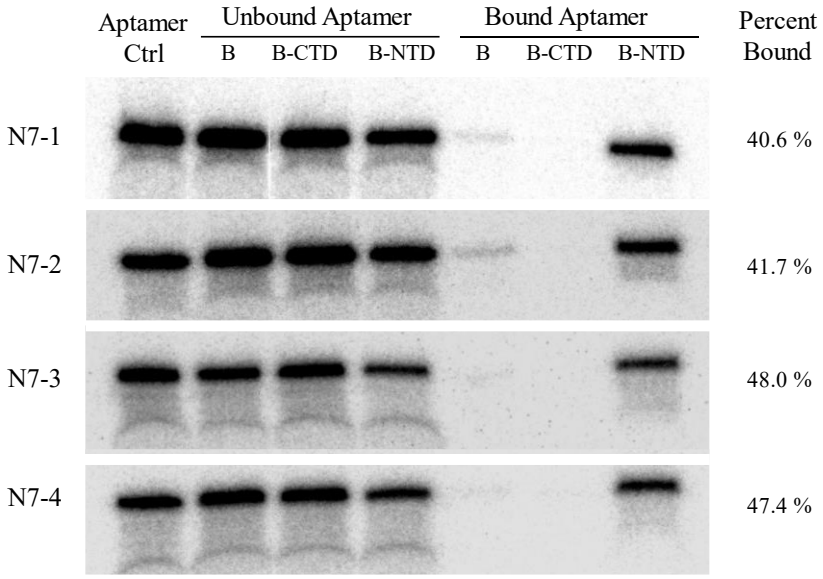


Figure 4.7 Binding test of N7 aptamers with a bead-based assay. Binding of N7 aptamers were tested against unloaded beads (B), bead-immobilized DnaK-CTD (B-CTD) and bead-immobilized DnaK-NTD (B-NTD). Aptamers (130 nM) were incubated with beads and protein (130 nM) for one hour. A magnetic rack was used to pellet the beads and the supernatant with unbound aptamers was removed. Beads were washed three times and resuspended in buffer (bound aptamers). All samples were run on denaturing urea PAGE (10 %). Gels were exposed to phosphor plates and imaged on Typhoon. Binding of aptamers was measured using ImageJ. Background binding to B ranged from 2.8 to 6.8 %. No significant binding of B-CTD was detected. Binding of aptamers to B-NTD was measured to be 40.6 % for N7-1, 41.7 % for N7-2, 48.0 % for N7-3, and 47.4 % for N7-4.

4.6. Characterization of binding affinity of N7-1 and N7-2

Binding of aptamers N7-1 and N7-2 to DnaK-NTD was assessed using the dot-blot assay. Protein was immobilized on magnetic beads and incubated with the aptamers. The mixture was separated by spotting the sample onto a nitrocellulose membrane; beads with protein and bound aptamers stayed on the nitrocellulose membrane, while unbound aptamers flowed through to the nylon membrane. The nitrocellulose and the nylon membranes were dried and exposed to storage phosphor plates, which were imaged. Binding of aptamers to the target was measured using ImageJ. Both aptamers N7-1 and N7-2 were tested against beads, B-CTD and B-NTD. N7-1 bound to B- NTD in a concentration-dependent manner with significant binding detected from 50 nM to 1000 nM **(Figure 4.8)**. N7-1 showed background binding to beads at the highest concentration tested; however, the binding was not concentration-dependent, and it is removed by optimization **(Figure 4.8)**. Similarly, N7-2 bound to B-NTD between 50 to 1000 nM with increasing binding at higher concentrations. N7-2 bound weakly to the beads only at high concentrations **(Figure 4.8)**. Neither N7-1 nor N7-2 showed non-specific binding to B-CTD **(Figure 4.8)**. Binding of aptamers N7-1 and N7-2 to B-NTD was quantified and used to create binding curves using GraphPad Prism non-linear regression for ligands with one-binding site (N=1); N7-1 had a Kd of 59.97 nM (R^2 0.9942), and N7-2 had a Kd of 114.1 nM (R^2 0.9987) **(Figure 4.9)**.

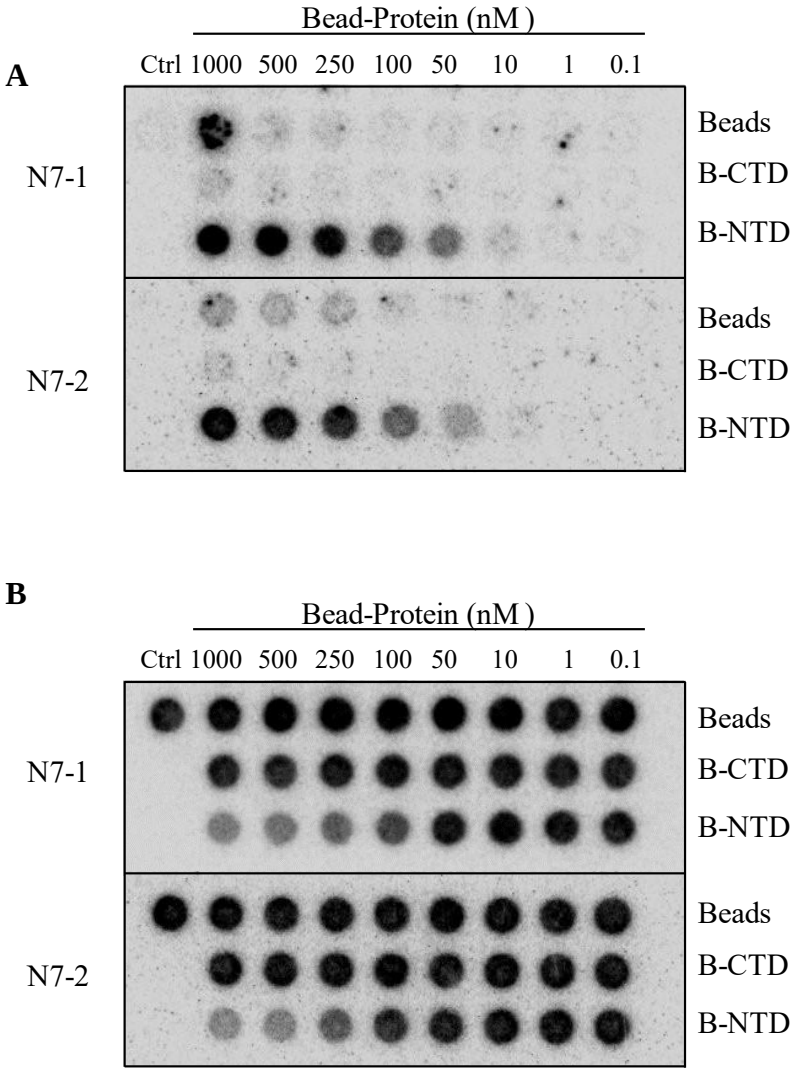


Figure 4.8 Dot-blot assay for aptamers N7-1 and N7-2. Radioactive aptamers N7-1 and N7-2 were incubated with various concentrations of beads, bead-immobilized DnaK-CTD (B-CTD), and bead-immobilized DnaK-NTD (B-NTD) (0.1 to 1000 nM). Loading control contained equivalent amount of aptamer in buffer. **(A)** Samples were applied onto nitrocellulose membrane, beads, B-CTD and B-NTD with bound aptamer remained. **(B)** Unbound aptamers flowed through onto the nylon membrane. Membranes were dried and exposed to phosphor plates for imaging with Typhoon. Binding of aptamers was calculated using ImageJ by calculating intensity of aptamer on nitrocellulose versus nylon membranes.

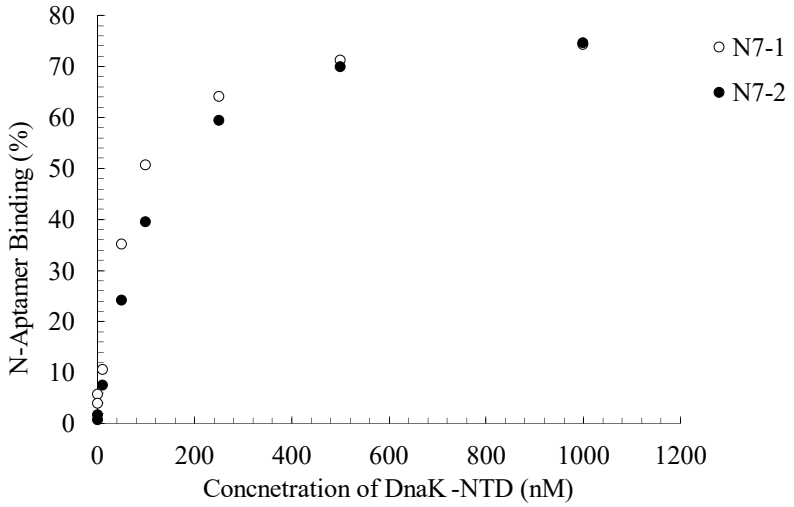


Figure 4.9 Binding curves of aptamers N7-1 and N7-2 to *F. nucleatum* DnaK-NTD. Binding of aptamers to DnaK-NTD was tested using a dot-blot assay and aptamer binding was calculated as a percent. Aptamer binding curves were created using GraphPad Prism non-linear regression for ligands with one-binding site. Dissociation constant (Kd) of aptamers was calculated to be 59.97 nM (R^2 0.9942) for N7-1, and 114.1 nM (R^2 0.9987) for N7-2.

4.7. Discussion and conclusion

The initial attempt to isolate aptamers for *F. nucleatum* DnaK-NTD failed due to DNA cross-contamination. Deep sequencing data suggested that the aptamer candidates (NAp 1, NAp2, NAp 3 and NAp 4) were positively selected over the selection process. However, only aptamer NAp 1 showed significant binding to

DnaK- NTD and it had very low binding affinity towards its target (Kd estimate 3 to 13 µM). Previously the same library and primer stocks were used for aptamer selection against *C. difficile* DnaK-NTD. Upon further inspection, it was discovered that the top aptamers produced by this selection were the same aptamers found by the selection against *C. difficile* DnaK-NTD. The libraries for both selections had such an extensive overlap in sequences that the *F. nucleatum* selection did not yield any unique classes of sequences. Analysis of the sequence composition at rounds 2, 4 and 6 from the *F. nucleatum* DnaK-NTD selection show that contaminating sequences were present from the beginning of the experiments. This enrichment pattern suggests that the emergence of the aptamer candidates was influenced by their abundance at the beginning of selection. Thus, enrichment of DNA sequences was not only driven by selective pressure towards binding sequences but also by an initial copy number bias.

Undeniably, the initial aptamer selection was influenced by the contamination of DNA sequences previously selected. However, it is important to note that selection was conducted while following strict guidelines to avoid DNA contamination. Remarkably, there was no evidence of contaminating DNA when conducting PCR reactions; that is, the negative control (no template) did not amplify. This evidence, along with the sequencing results, suggests that the initial library used was biased. Hence, the handling protocols for purification, storage and usage of libraries implemented in the lab should be revised. Additionally, it is likely that the structural similarity between the targets lead to the perpetuation of

contaminating sequences in the DNA pool. After all, DnaK from *C. difficile* and *F. nucleatum* have 66.23 % overlap in amino acid sequence identity. Therefore, extra caution should be taken when devising experimental designs for similar targets to avoid contamination.

The evidence presented in this chapter suggests that the second aptamer selection process was successful in isolating sequences that bind *F. nucleatum* DnaK-NTD. First, the sequencing data shows an initially naïve DNA pool that became progressively enriched over each selection cycle; substantial enrichment occurs at NR6 after introduction of counter-selection and increasing the selection pressure. Second, all four aptamer candidates bound B-NTD specifically and not B-CTD or the free beads. Moreover, further examination of the binding of two aptamers candidates showed that they bound B-NTD with suitable affinity (60 and 114 nM). As a result, we can conclude that the selection strategy was successful in identifying binding sequences.

While supporting evidence is presented regarding the affinity of the selected aptamers towards B-NTD, they did not appear to bind DnaK-NTD in solution as shown by the EMSA experiments. Since the structure of DnaK is highly dynamic and depends on its binding partners (Matthias P. Mayer & Gierasch, 2019), it is reasonable to hypothesize that the aptamers isolated may be able to bind only a specific subset of conformations adopted by DnaK-NTD. Perhaps the conformation of free DnaK-NTD compared to immobilized B-NTD could be different enough to override the aptamers' molecular recognition mechanisms.

Nonetheless, the conformation recognized by these aptamers could still be biologically relevant and further testing against *in vitro* and *in vivo* DnaK would be needed. On the other hand, the failure of EMSA may also be caused by the conditions of the test. In this protocol, the protein-aptamer mixture is incubated in reaction buffer, then applied to the gel and separated by electrophoresis. The gel running buffer and gel conditions are different to the reaction buffer (ion concentration, pH, etc.); thus, it is possible that the aptamers can recognize DnaK-NTD, but their interaction is weak under the latter conditions and they dissociate. Albeit, different gel running buffers and reaction buffers were tested to improve the assay but none of them worked.

The four aptamer candidates identified in this selection bind to *F. nucleatum* DnaK-NTD and the affinity of two aptamers was characterized as promising for subsequent work. However, these aptamers only recognize bead-immobilized DnaK-NTD and not to the protein in solution. These findings highlight the limitation of bead-based selection where ligands become biased towards protein conformations adopted when the protein is immobilized. Furthermore, they also highlight the molecular complexity of DnaK as a highly dynamic protein target able to adopt multiple conformations. More evidence needs to be gathered to conclude whether these aptamers can be used for downstream applications.

CHAPTER 5:APTAMER SELECTION FOR *F. NUCLEATUM* DNAK C-TERMINAL DOMAIN

5.1. Selection of aptamers for DnaK-CTD

Aptamer selection for DnaK-CTD was performed for ten rounds using the ssDNA library FN40R-7 and bead-based SELEX. While this library was previously used to isolate aptamers for *C. difficile* DnaK-CTD, several safeguards were employed to avoid DNA cross-contamination between selections. Composition of DNA pools were analysed with Illumina deep sequencing. Round five to seven were sequenced to monitor the progression of selection. Then, rounds eight to ten were sequenced at the end of selection.

Individual sequences and clusters were ranked by their abundance in each DNA pool, a summary of the sequencing data is presented in **Table 5.1**. The first set of data, CR5 to CR7, suggested mild enrichment of the DNA pool. CR7 showed significant enrichment of the top cluster, which had 2408 reads, while all other clusters in this pool had less than 100 reads. As such, the top sequence at CR7 was considered as a potential aptamer candidate and termed C7-1. The data from CR8 to CR10 shows strong evidence of enrichment as the average frequency of clusters progressively increased and the number of unique clusters steadily decreased **(Table 5.1)**. Furthermore, significant enrichment of the top cluster is evident between round 8 and 9 as the frequency of the top cluster jumped up from 0.4 % to 35.3 % **(Table 5.1)**. Overall, the sequencing data shows that selection leads to enrichment towards specific sequences.

Sequencing data from round CR10 was used to further identify aptamer candidates by looking at the top 10 cluster families **(Table 5.2)**. Cluster 1 dominated the pool, and Clusters 2 to 5 were significantly enriched and made up 33.54 % of the library. Clusters 6 to 10 were moderately enriched and made up only 1.39 % of the pool. The remainder of the sequences made up 9.58 % of the pool. Frequency of the top five clusters in CR9 and CR10 were examined to assess their enrichment patterns; all five clusters were found to increase in frequency **(Figure 5.1)**. Therefore, selection drove the composition of the DNA pool towards few selected sequences. The top five clusters at CR10 were considered DNA candidates and called C10-1, C10-2, C10-3, C10-4, and C10-5.

The DNA library from the *F. nucleatum* DnaK-CTD selection (F.n. CR10) was compared to the library *C. difficile* DnaK-CTD selection at round 10 (C.d. CR10) to check for cross-contamination of DNA. No significant overlap between the two pools was found. The top 20 clusters of C.d. CR10 are not found in F.n. CR10. Conversely, the top F.n. CR10 sequences had little overlap with the sequences in the C.d. CR10 pool: C10-1 was ranked 25, C10-3 was ranked 61, C10-4 was ranked 35, and C10-9 was ranked 75 in C.d. CR10. The remaining top clusters (2, 5, 6, 7, 8 and 10) were not present in C.d. CR10. Thus, the enrichment of sequences observed in F.n. CR10 is due to selective pressure of the selection process and not due to DNA contamination.

Table 5.1 Summary of sequencing data for CR7 to CR10. Sequencing data from DNA pools after five to ten rounds of selection against *F. nucleatum* DnaK-CTD. The number of sequences read by the equipment for each pool is represented as the total number of reads. The number of unique clusters per round was calculated by organizing the individual reads into unique clusters (families of similar sequences). Calculated average number of reads per cluster and the average frequency of clusters, and the frequency of the most abundant cluster in each round is presented.

Round	Total Number of Reads	Number of Unique Clusters	Average Number of Reads per Cluster	Average Frequency of Clusters (%)	Frequency of Top Cluster (%)
CR10	504931	46853	10.77	2.13×10^{-3}	55.5
CR9	565502	284323	1.989	3.52×10^{-4}	35.3
CR8	595354	589085	1.011	1.70×10^{-4}	0.40
CR7	488766	480001	1.018	2.09×10^{-4}	0.49
CR6	149630	148118	1.010	6.75×10^{-4}	2.00×10^{-3}
CR5	357001	353663	1.009	2.81×10^{-4}	8.34×10^{-4}

Table 5.2 Top 10 clusters at CR10. After ten rounds of selection against DnaK-CTD the DNA pool has become enriched. The top 10 families of sequences are presented, and their relative abundance is described as percentage.

Cluster	Abundance	Sequence
1	55.49	CATCAAAGCTAGTCTTCCGAAAACGTGCCGTAAGGTGTGC
2	12.71	GATACGGTCGCGTGAGACGACGTATGCCCAGCCGTTCACA
3	12.45	CCACCGCGAAAGACGGTCCCGAGCCCGCGAGTGGCACACA
4	7.36	TACACTTTGTCCGCCATAAGTGAATACGTTTGAGTACTAGC
5	1.03	TGACGCATTATGGTTTATATGCGCGGCCGTCTGAGGGGAC
6	0.41	ACTGTCCCCTCTTTTTAATAGCCGCGTATAATGTTTGTCC
7	0.39	TGGATTGGAGGATGTTTGCGCCGGTTCGAGGTCGCGTTTT
8	0.30	TGTTTGATAGAAGATGTTTCCGCGTTCGGTGCCGGAGGGT
9	0.18	TTTACTTGACGATCTTGTTTTATGTTCGCACCTGTCTGAA
10	0.11	TGCGCGGGTGGATATCTGGCCAGGGAAGGGGGTGGTGGTG

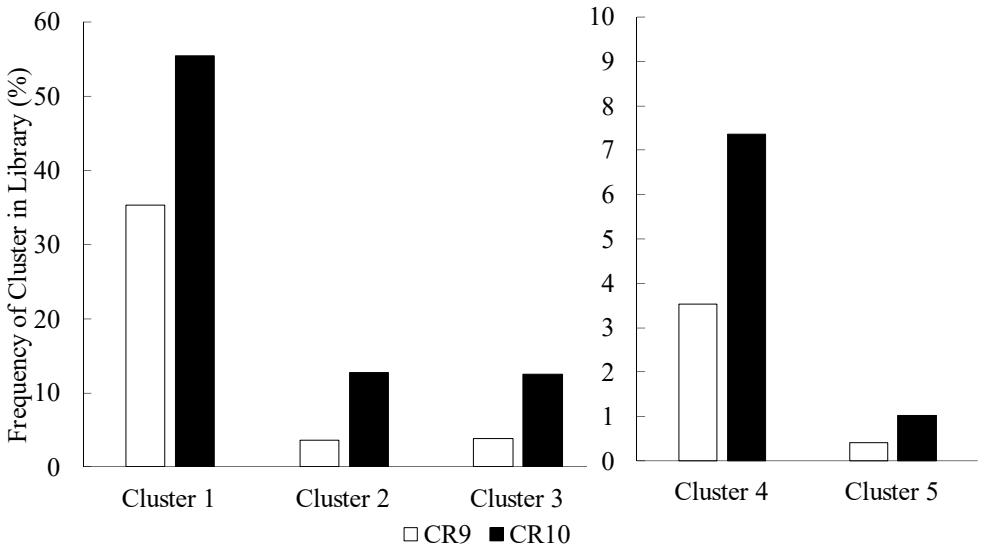


Figure 5.1 Abundance of top five sequence clusters in CR9 and CR10. The top five clusters at CR10 were identified and their frequencies were tracked to previous rounds of selection. The frequencies of top clusters are presented as a percentage of abundance in the pool. Clusters 1 to 5 significantly increased in abundance.

5.2. Characterization of binding of aptamer candidates against DnaK-CTD

Aptamer candidates were tested against DnaK-CTD using EMSA. In this method, the radiolabelled aptamers were each incubated with different concentrations of protein in solution, then, the mixtures were run on native PAGE. Binding of aptamer to the protein was detected as an upward shift of the DNA band. Aptamers C7-1, C10-1, C10-2, C10-3, C10-4, and C10-5 were all able to bind to DnaK-CTD with increasing binding detected at higher concentrations of protein **(Figure 5.2)**. However, the gels produced from these EMSAs often were smeared and the bands were not clear. As a result, binding of the aptamers could not be quantified. Extensive optimization of the gel conditions and the binding reaction was conducted to improve the quality of the EMSA gels. The binding buffer was also modified by testing different concentrations of monovalent and divalent cations. Notably, incubation of DnaK-CTD with aptamers in selection buffer without calcium (Ca^{2+}) lead to the degradation of the aptamers **(Supplemental Figure 4)**. Nonetheless, the optimization of the protocol led to a dramatic improvement on the quality of the EMSAs **(Supplemental Figure 5)**.

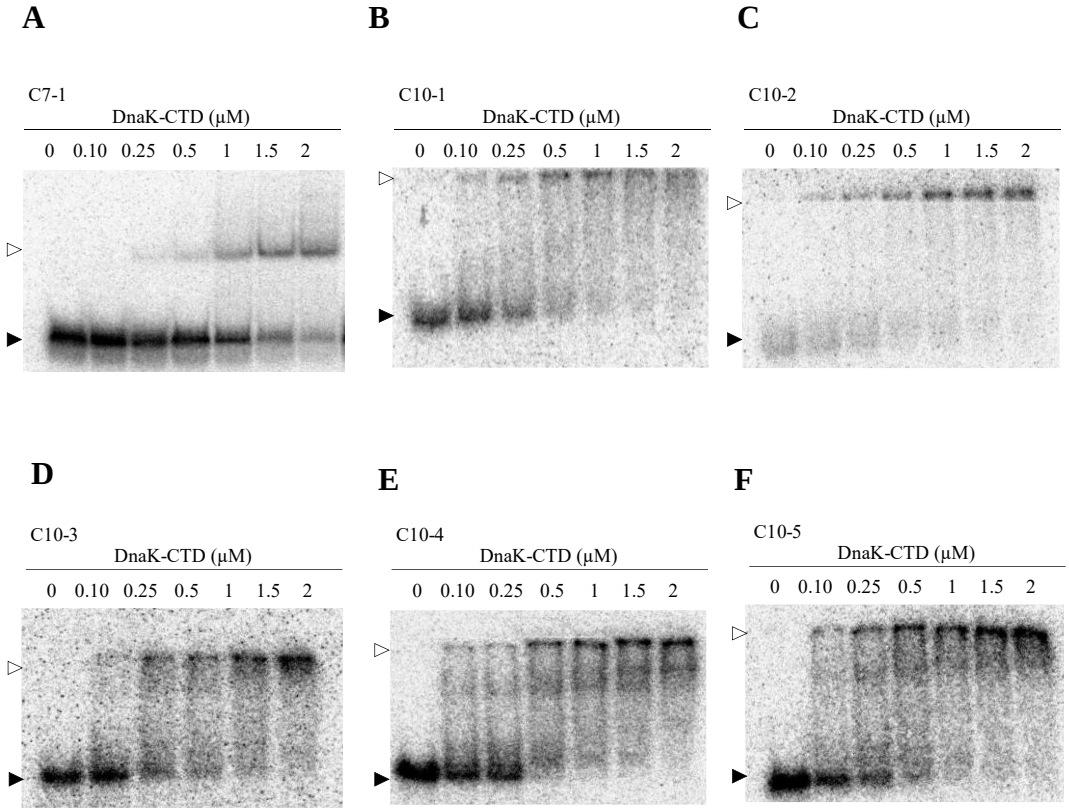


Figure 5.2 Binding test of C-aptamers to *F. nucleatum* DnaK-CTD. EMSA was performed with aptamers **(A)** C7-1, **(B)** C10-1, **(C)** C10-2, **(D)** C10-3, **(E)** C10-4, and **(F)** C10-5 and various concentration of DnaK-CTD protein (0 μM to 2 μM). Aptamers were radiolabeled, incubated with target for one hour, and run on native PAGE (7 %). Gels were exposed to phosphor plates for and scanned with Typhoon. The black arrowhead (►) indicates the position of the free aptamer, while the white arrowhead (▷) shows the position of the protein-aptamer complex.

5.3. Troubleshooting binding assays: Degradation of DNA

The contaminant causing the degradation of the aptamers was found in the DnaK-CTD protein stock. To assess the identity of this contaminant, two DnaK-CTD stocks were tested; Aug2019 and May2019. Aptamers C7-1, C10-1, N7-1, and N7-2 were incubated for with 1 µM of DnaK-CTD or with buffer, then, the mixture was run on a native PAGE and on a denaturing urea PAGE **(Figure 5.3)**. The native gels show that both stocks contain a protein that bind the aptamers, which is seen as an upward shift of the DNA band **(Figure 5.4 A)**. Both C-aptamers and N-aptamers bind a protein in solution; though, the C-aptamers had a wider shifted band that may suggest DNA binding to both DnaK-CTD and the contaminant **(Figure 5.4 A)**. The denaturing gels showed clear degradation of all aptamers in the presence of protein **(Figure 5.4 B)**. Digestion of the aptamers created a smearing pattern, which suggests that digestion occurs on a 3' to 5' as the DNA is labeled on the 5' end **(Figure 5.4 B)**.

DnaK-CTD protein was previously found to be highly pure; they showed one band at the expected size **(Figure 3.4)**. DnaK-CTD stock were evaluated for protein contaminants by separation on 15 % and 20 % SDS PAGE gels. The main band of protein in the stock split slightly in the 15 % SDS PAGE, and more significantly in the 20 % SDS PAGE **(Supplemental Figure 6)**. These results revealed the presence of another protein in the stock that has a slightly lower mass than DnaK-CTD. All together, the DnaK-CTD stock seems to be contaminated

with an unknown protein with a molecular weight between 25 and 29 kDa that has a 3' to 5' ssDNA nuclease activity and it is inhibited by the presence of Ca^{2+}.

The bead-loading protocol was examined to assess whether selection was conducted in the presence of the contaminant. During selection, magnetic beads are incubated with the DnaK-CTD stock. Then, the loaded beads are aggregated with a magnetic rack and the initial supernatant (Si) is removed. Beads are washed three times by resuspending them in buffer, then, the beads are aggregated, and the buffer is removed. Lastly, the beads are resuspended in a known concentration. Bead-immobilized DnaK-CTD (B-CTD) was incubated for one hour with aptamer C10-1. Samples were run on a denaturing PAGE. The aptamer was not digested by B-CTD **(Supplemental Figure 7)**. To further examine this protocol, the bead-loading protocol was repeated with five wash steps **(Figure 5.5 A)**. Samples for each wash were kept for the bead fractions (Bead-CTD) and the supernatant fractions. Each sample was incubated with ssDNA for one hour and run on a denaturing PAGE. The initial supernatant contained most of the nuclease activity, which was observed as significant digestion of the aptamer **(Figure 5.5 B)**. Some minor smearing is visible in the B-CTD fraction after one to three washes, and in the four and five supernatant fractions **(Figure 5.5 B)**. Thus, the contaminant is completely removed from the bead fraction after five washes with buffer.

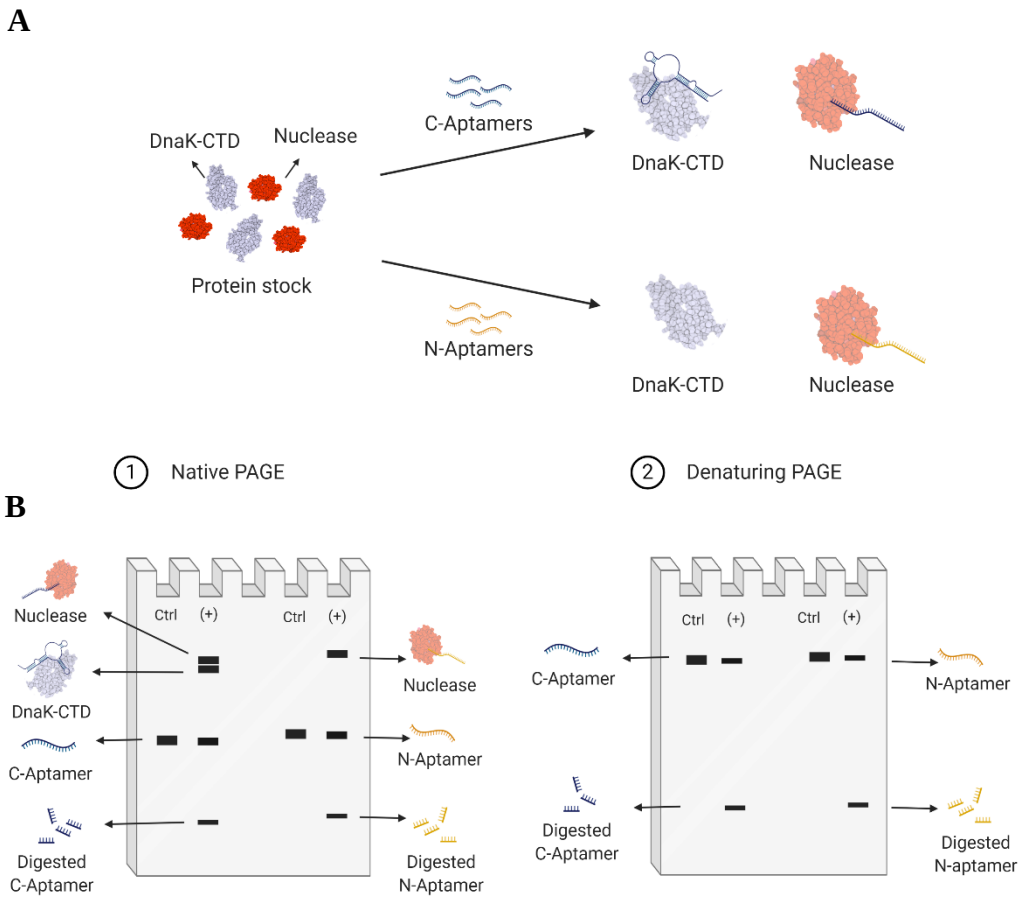


Figure 5.3 Schematic of aptamer degradation experiment. (A) The DnaK-CTD protein stock was incubated with C-Aptamers and N-Aptamers separately. In theory, C-Aptamers can be bound to DnaK-CTD and a contaminant with nuclease activity, while N-Aptamers only bind to the contaminant. **(B)** Samples are run on a native PAGE gel to detect the complexes formed by the aptamers; an upward shift of the DNA indicates an aptamer-protein complex and a downward shift indicates DNA degradation. The samples are run on a denaturing PAGE gel to separate the DNA by size; short fragments at the bottom of the gel indicated aptamer degradation.

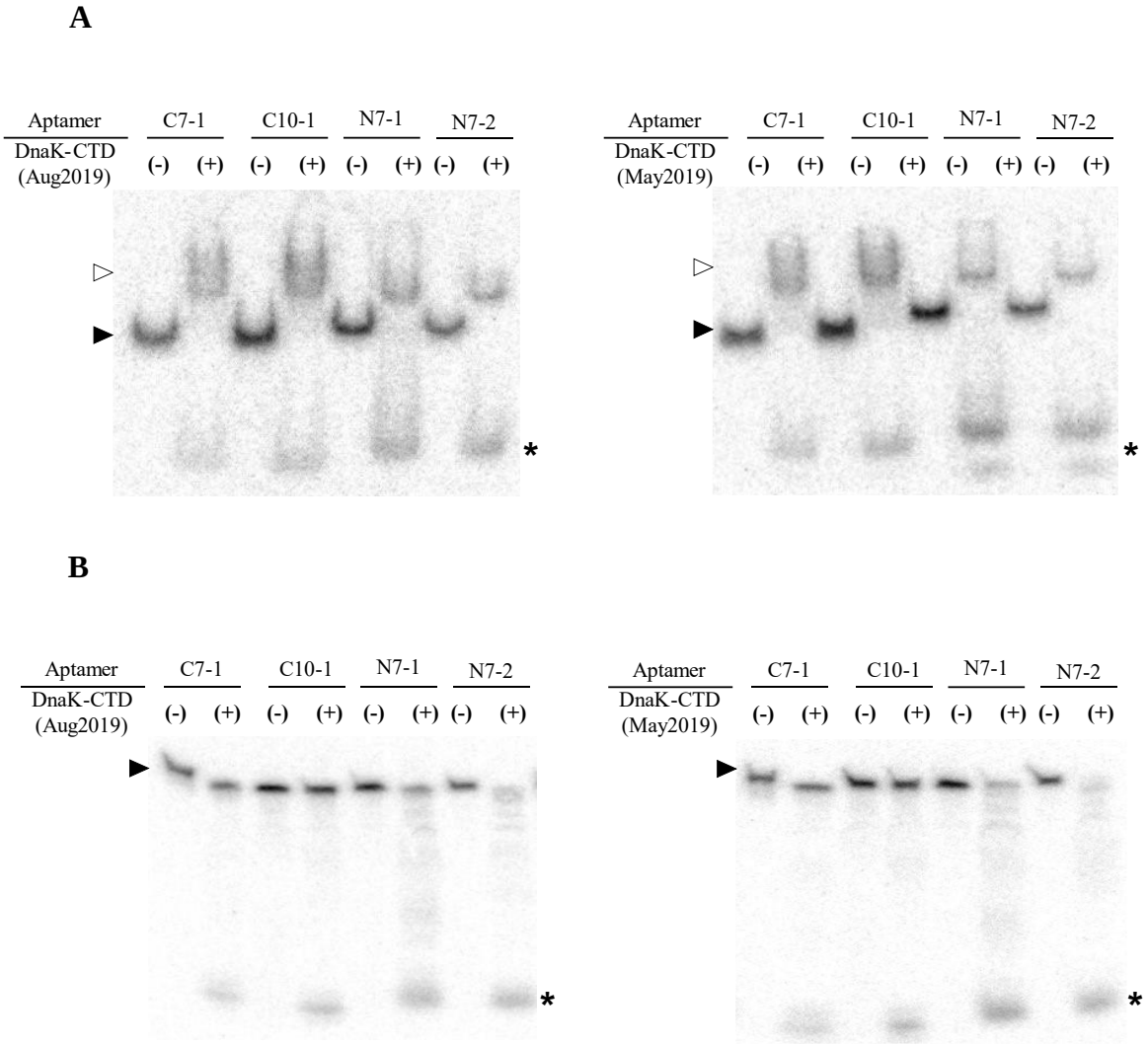
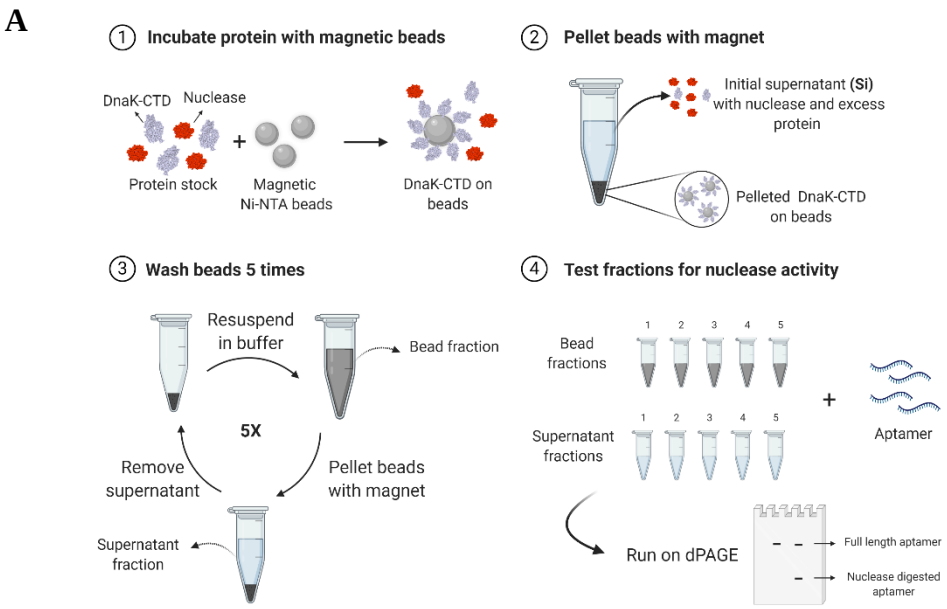


Figure 5.4 Nuclease contaminant test of DnaK-CTD. Aptamers C7-1, C10-1, N7-1, and N7-2 were incubated in buffer (-) or with 1 µM DnaK-CTD stocks (+) from August 2019, and May 2019. **(A)** Samples were run on a native PAGE (7%). Aptamers incubated with DnaK-CTD shifted upwards (white arrowhead), compared to the free aptamer (black arrowhead). Aptamer degradation is indicated by the asterisk (*). **(B)** Samples were also run on a denaturing urea PAGE (10%) to separate DNA by size. Full length aptamers are indicated by the black arrowhead. Aptamers treated with DnaK-CTD were degraded as shown by the smearing across the gel and the short fragments at the bottom of the gel (*).



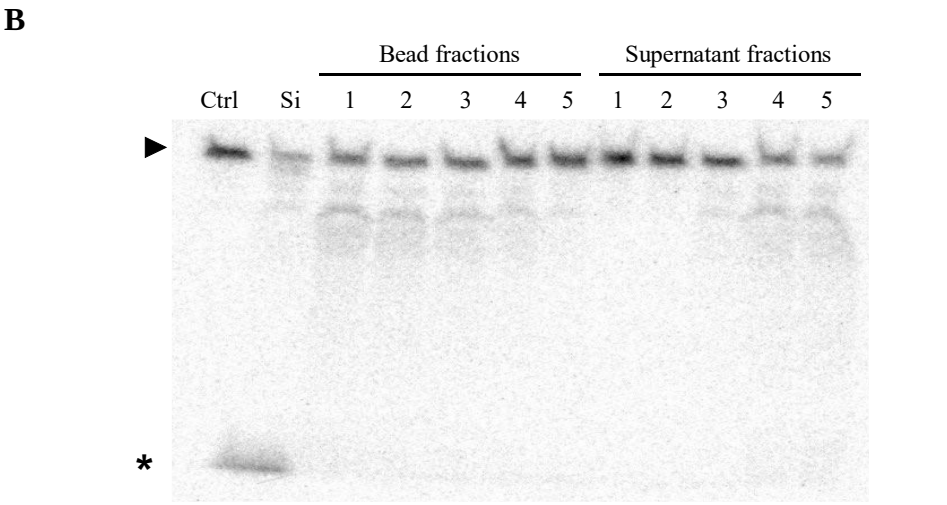


Figure 5.5 Nuclease activity is removed by loading DnaK-CTD onto beads. **(A)** Schematic of experimental overview. (1) DnaK-CTD stock was incubated with magnetic beads. (2) Beads were pelleted with a magnetic rack and the supernatant was removed (Si). (3) Beads were washed five times by resuspending in buffer, then the beads were aggregated, and the supernatant was removed. (4) The bead and supernatant fractions for washes 1 to 5 were incubated with aptamer for 1h and ran on a dPAGE (10%). **(B)** Gel shows aptamer (►) for control, initial supernatant and bead and supernatant fractions. Degradation of the aptamer is indicated by asterisk (*). The numbers (1-5) indicate the number of washes for a given sample.

5.4. Calculation of binding affinity of C-aptamers

Binding affinity of C-aptamers towards DnaK-CTD was tested using dot-blot assays. The proteins were immobilized on magnetic beads and thoroughly washed with selection buffer to remove the contaminant. B-CTD were diluted to a gradient of concentrations and incubated with radiolabeled aptamers. The dot-blot apparatus was set up with a nitrocellulose (NC) and a nylon membrane to separate bound and unbound aptamers, respectively **(Figure 2.2 B)**. Aptamers were tested against B-CTD, B-NTD and free beads **(Figure 5.7)**. C-aptamers were found to bind B-CTD in a concentration-dependent manner, and to not bind the B-NTD and bead controls. Binding of the C-aptamers was quantified, and the data (N=3) was used to analyze binding kinetics with GraphPad prism non-linear fit with one-site binding. Binding curves for C7-1, C10-1, C10-2, C10-3, C10-4, and C10-5 were graphed **(Figure 5.7)**. The Kd of the C-aptamers were calculated to be 202 ± 75.83 nM (R^2 0.81) for C7-1, 212.1 ± 68.52 nM (R^2 0.84) for C10-1, 263.5 ± 46.48 nM (R^2 0.96) for C10-2, 270 ± 49.75 nM (R^2 0.95) for C10-3, 279.4 ± 43.5 nM (R^2 0.97) for C10-4, and 273.5 ± 49.45 nM (R^2 0.96) for C10-5.

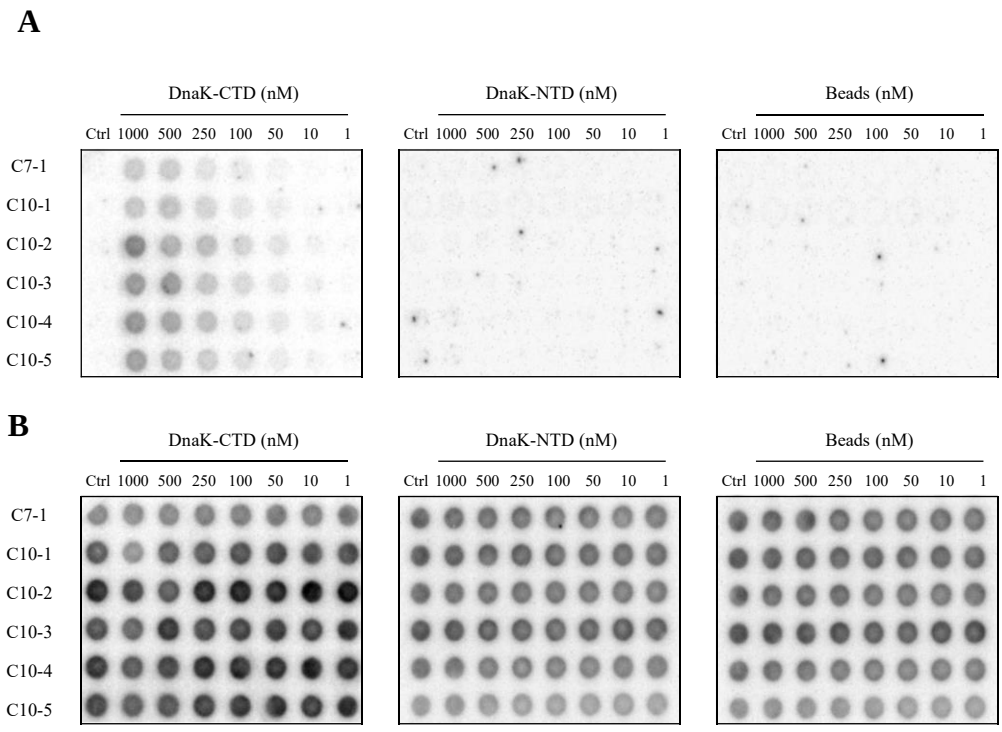


Figure 5.6 Dot-blot assay for C- aptamers. Radioactive aptamers C7-1, C10-1, C10-2, C10-3, C10-4, and C10-5 were incubated with various concentrations of bead-immobilized DnaK-CTD (B-CTD), bead-immobilized DnaK-NTD (B-NTD) and beads (1000 to 1 nM). Loading control contained equivalent amount of aptamer in buffer. **(A)** Samples were applied onto nitrocellulose membrane, beads, B-CTD and B-NTD with bound aptamer remained. **(B)** Unbound aptamers flowed through onto the nylon membrane. Membranes were dried and exposed to phosphor plates for imaging with Typhoon. Binding of aptamers was calculated using ImageJ by calculating intensity of aptamer on nitrocellulose versus nylon membranes.

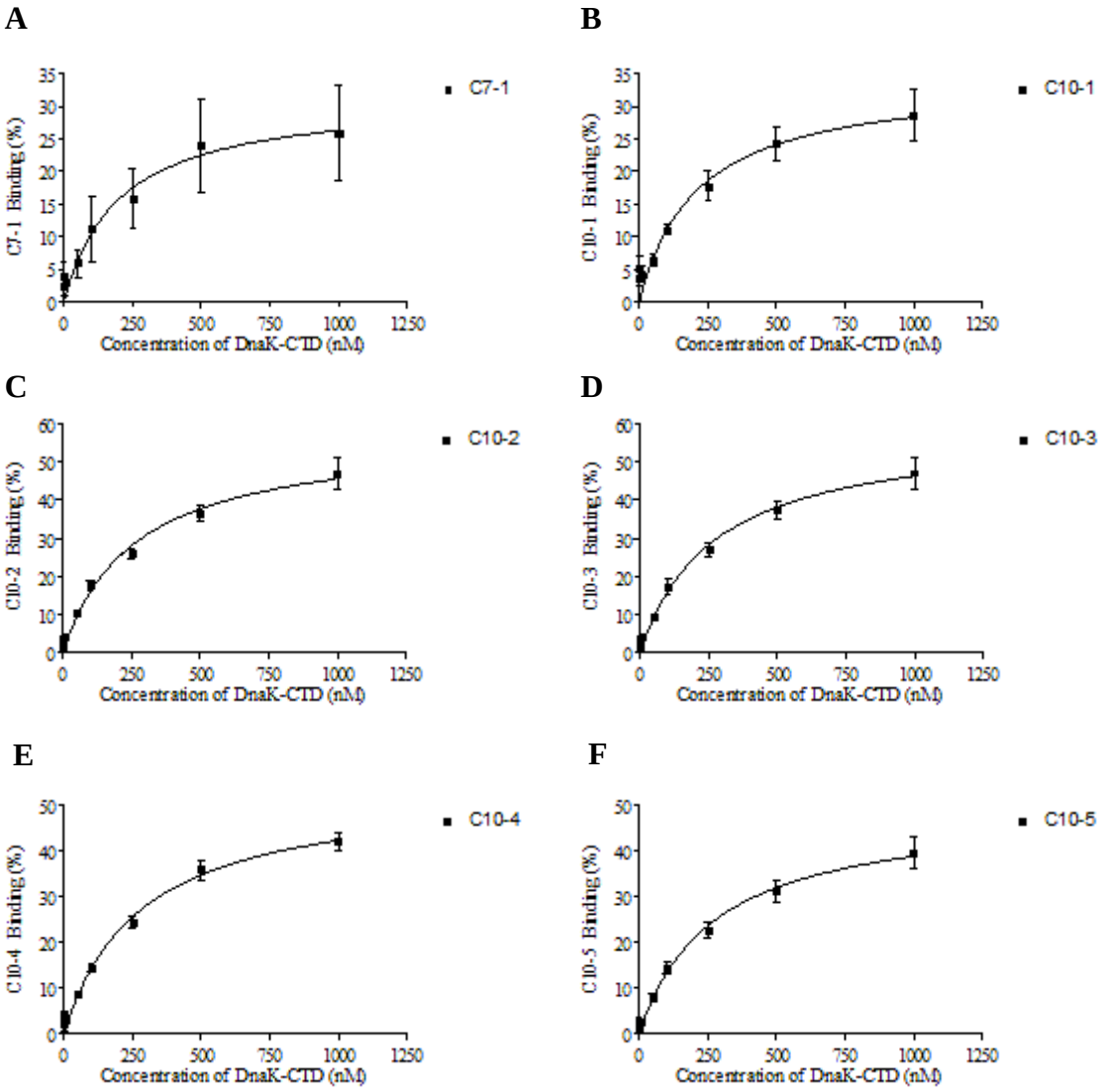


Figure 5.7 Binding curves of C-aptamers with *F. nucleatum* DnaK-CTD. Binding of aptamers to DnaK-CTD was tested using a dot-blot assay in triplicate and aptamer binding was calculated as a percent. Aptamer binding curves were created using GraphPad Prism non-linear regression for ligands with one-binding site and dissociation constants (Kd) of aptamers were calculated. **(A)** C7-1 Kd of 202 ± 75.83 nM (R^2 0.81), **(B)** C10-1 Kd of 212.1 ± 68.52 nM (R^2 0.84), **(C)** C10-2 Kd of 263.5 ± 46.48 nM (R^2 0.96), **(D)** C10-3 Kd of 270 ± 49.75 nM (R^2 0.95), **(E)** C10-4 Kd of 279.4 ± 43.5 nM (R^2 0.97), **(F)** C10-5 Kd of 273.5 ± 49.45 nM (R^2 0.96).

5.5. Discussion and conclusion

The selection of aptamers for *F. nucleatum* DnaK-CTD proceeded effectively despite multiple problems with the methodology used as significant problems were encountered with both the ssDNA library and the protein stock. The library used for this selection was the same library used for the *C. difficile* DnaK-CTD. As previously discussed in chapter 4, using the same libraries for similar targets increases the risk of DNA cross-contamination of DNA. The evidence presented in this chapter suggests that the library for *F. nucleatum* DnaK-CTD selection did not become contaminated. Additionally, the top aptamers from C.d. R10 are not present in the *F. nucleatum* selection, and the enrichment pattern observed does not suggest copy number bias. This positive outcome may be because a new stock of library and primers were ordered for this project, as a result, all stocks were purified and prepared without the influence of the *C. difficile* aptamers. Additionally, cautionary practices were introduced at round four to avoid any additional sources of cross-contamination. Therefore, the library for the *F. nucleatum* selection was not contaminated due to the implementation of precautionary practices and these handling protocols should be used in future selection work in the lab to avoid cross-contamination of DNA.

Contamination of this selection occurred instead at the protein level. Isolated proteins are examined using the SDS gel method as a standard practice in the biological sciences, here, the presence of one band at the expected size is

considered a successful purification. However, this standard failed in the case of the purified DnaK-CTD. All the isolated batches of DnaK-CTD showed a singular band at the appropriate size when the samples were run on 7% SDS PAGE. Yet, the stocks of DnaK-CTD had nuclease activity. Further investigation showed that the DnaK-CTD band split when run on high percentage SDS PAGE **(Supplemental Figure 6)**. The stocks digested DNA in 3' to 5' direction, and the nuclease activity was inhibited by the presence of Ca^{2+} in solution. When both N- and C-aptamers were tested for nuclease degradation, it was shown that the N-aptamers were degraded more effectively than the C-aptamers **(Figure 5.4)**. This may be evidence that the nuclease contaminant is biased towards ssDNA, as the 3' and 5' ends of the C-aptamers base pair and are functionally double stranded at the ends. On the other hand, it may suggest that the C-aptamers are somewhat protected from degradation by binding DnaK-CTD. Nonetheless, the contaminant nuclease could be removed using Ni-NTA beads **(Figure 5.5)**. Therefore, the nuclease contaminant was inadvertently removed by the bead-based SELEX protocol, and the selection was performed towards the right protein target.

Searching the Uniprot database for *E. coli* nucleases between 29 to 25 kDa with 3' to 5' yielded two potential candidates: TatD exonuclease (28.974 kDa) and Exodeoxyribonuclease 10 (25.133 kDa). TatD is most likely to be the contaminant of the two nucleases. It is only 376 Da smaller than DnaK-CTD, and it has four surface histidine residues that could potentially bind the His-column weakly. Furthermore, it has a PI of 5.18 while DnaK-CTD has PI of 5.15, as such, they are

likely to elute from the anion exchange column at similar concentrations of NaCl. Interestingly, TatD is also reported to be inhibited by Ca^{2+} and to have preference towards ssDNA (Chen et al., 2014). As such, these data indicate that TatD is the likely contaminant present in the DnaK-CTD protein stock.

Characterization of the binding properties of the C-aptamers was done using EMSA and dot-blot assays. The EMSAs were inconclusive due to the presence of contaminating nuclease degrading the aptamers. The dot-blot assay showed that the C-aptamers bound DnaK-CTD in a concentration-dependant manner and ranged in affinity from 202 to 279 nM. While it was demonstrated that the B-CTD used in the dot-blots did not have nuclease activity **(Supplemental Figure 7)**, it is still possible, although unlikely, that the binding observed is not caused by the aptamers binding B-CTD but due to the contaminant. Nevertheless, the collection of evidence presented in this chapter suggests that the selection of aptamers toward *F. nucleatum* DnaK-CTD was successful and that the aptamers isolated have suitable affinity towards their target.

CHAPTER 6:APTAMER RESELECTION AGAINST *F. NUCLEATUM* DNAK

6.1. Characterization of N-Aptamers and C-Aptamers against DnaK

Aptamers C10-1, C10-2, C10-3 and N7-1, N7-2 and N7-3 were assessed against DnaK using dot-blot assay. Aptamers were incubated with various concentrations of bead-immobilized DnaK (B-DnaK) for one hour. Samples were applied onto the nitrocellulose membrane on the dot-blot apparatus and washed twice. Unbound aptamers flowed through an were capture on the nylon membrane **(Figure 2.2 B)**. Two different batches of DnaK were tested against the C-aptamers and N-Aptamer from February 2020, DnaK (Feb20), and DnaK from April 2018, DnaK (Ap18). No significant binding of aptamers to either DnaK (Feb20) nor DnaK (Ap18) was observed under these conditions **(Figure 6.1**).

6.2. Overview of reselection protocol

Given that the selected aptamers did not bind the whole protein, reselection was conducted using the full-length DnaK protein. The reselection experiment consisted of three rounds of selection with DnaK with primed libraries previously used for selection against DnaK-NTD (FN40R-10, N-library) and DnaK-CTD (FN40R-6, C-library). Reselection used libraries that are biased towards the target of interest, DnaK-NTD or DnaK-CTD, but that still have a high degree of diversity such that different sequences can emerge from the pool. To this end, the N-library at round 5 and the C-library at round 7 were chosen; they both have an average read per cluster of 1.01 and the top sequence made up ~0.45 % of the

pool **(Table 4.3, Table 5.1)**. Thus, these slightly enriched but still had highly diverse libraries were selected.

Three rounds of selection were conducted against DnaK using bead-based selection with the N-library and the C-library. In each round of reselection 100 pmol of library was used, while the amount of DnaK target was 130 nM in the first round, 40 nM in the second round, and 20 nM in the third round. Rounds of reselection with the N-library were called RS-N6, RS-N7 and RS-N8; while rounds with the C-library were called RS-C8, RS-C9 and RS-C10. Rounds RS-N7, RS-N8, RS-C9 and RS-C10 were analysed with Illumina deep sequencing.

6.3. N-Library reselection of aptamers with DnaK

Sequencing data from RS-N7 and RS-N8 were organized into clusters of similar sequences and ranked by their relative abundance in the DNA pool. A summary of the sequencing data from RS-N7 had RS-N8 is presented in **Table 6.1**. The number of unique clusters increased, and the frequency of the top cluster decreased between rounds RS-N7 and RS-N8 **(Table 6.1)**. Thus, the DNA pool did not enrich over reselection.

The top clusters at RS-N7 and RS-N8 were the same top sequences found in the original selection, albeit they had much lower abundance. Since NR5 is the library used at the beginning of the reselection process it provides a useful baseline. The five N-aptamer candidates were tracked to NR5 and to the reselection libraries.

The abundance of clusters one to four decreased over reselection, and cluster five increased slightly **(Figure 6.2 A)**. The change in abundance of the top sequences between RS-N7 and RS-N8 was calculated as an enrichment factor; clusters one to four had large negative enrichment factors (30 to 80 times lower abundance), while cluster five increased by a factor of 25 **(Figure 6.2 B)**. Overall, these sequencing data suggests that the DNA pool is not enriched after three rounds of reselection.

Table 6.1 Summary of sequencing data for RS-N7 and RS-N8. Sequencing data from N-libraries after 2 and 3 rounds of reselection with DnaK, RS-N7 and RS-N8, respectively. The number of sequences read by the equipment for each library is represented as the total number of reads. The number of unique clusters per round was calculated by organizing the individual reads into unique clusters (families of similar sequences). Calculated average number of reads per cluster and the average frequency of clusters, and the frequency of the most abundant cluster in each round is presented.

Round	Total Number of Reads	Number of Unique Clusters	Average Number of Reads per Cluster	Average Frequency of Clusters (%)	Frequency of Top Cluster (%)
RS-N8	408587	405562	1.01	2.47×10^{-4}	0.23
RS-N7	408689	405404	1.01	2.47×10^{-4}	0.34

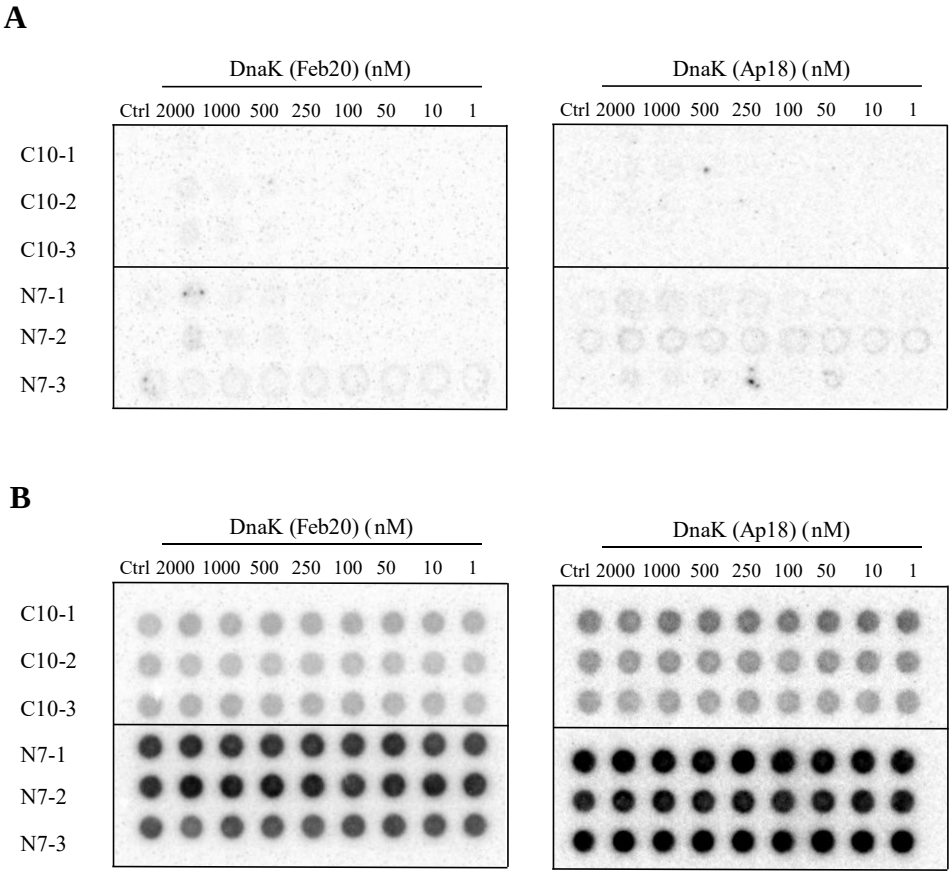


Figure 6.1 Dot-blot assay for C-aptamers and N-aptamers with *F. nucleatum* DnaK. Radioactive aptamers C10-1, C10-2, C10-3, N7-1, N7-2, and N7-3 were incubated with various concentrations of bead-immobilized DnaK (B-DnaK) (1 to 2000 nM). Two batches of DnaK were tested from February 2020 (left) and April 2018 (right). Loading control contained equivalent amount of aptamer in buffer. **(A)** Samples were applied onto nitrocellulose membrane B-DnaK with bound aptamer remained. **(B)** Unbound aptamers flowed through onto the nylon membrane. Membranes were dried and exposed to phosphor plates for imaging with Typhoon. Binding of aptamers was calculated using ImageJ by calculating intensity of aptamer on nitrocellulose versus nylon membranes. Aptamers do not bind significantly to B-DnaK.

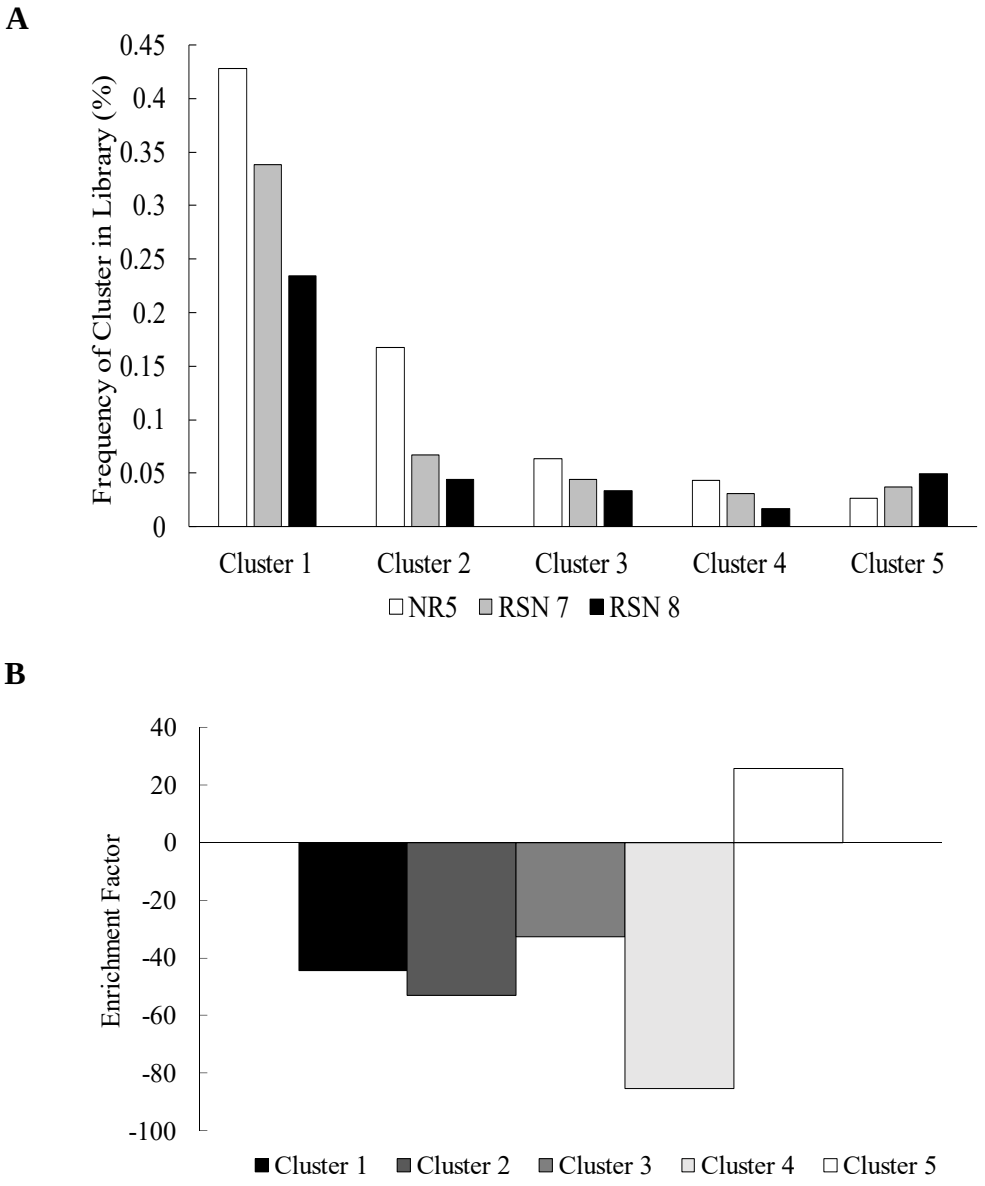


Figure 6.2 Effects of three rounds of reselection with DnaK on N-libraries. The top five clusters after selection with DnaK-NTD were identified and their frequency was tracked to the NR5 library and to the reselection libraries RS-N7 and RS-N8. **(A)** The frequency of top clusters is presented as percent abundance in library. **(B)** The enrichment factor of the clusters 1 to 5 between RS-N7 and RS-N8 was calculated and plotted.

6.4. C-Library reselection of aptamers with DnaK

Sequencing data for the C-library reselection was organized based on the relative abundance of the individual sequences and clusters. A summary of the sequencing data for the reselection libraries RS-C10 and RS-C9 is presented in **Table 6.2**. The libraries show signs of enrichment as the number of unique clusters decreased between RS-C9 and RS-C10, and the average frequency of clusters increased **(Table 6.2)**. Moreover, the top cluster is reported to be highly enriched at ~ 17 % compared to 0.49 % at the start of reselection (CR7).

The top 10 clusters in RS-C10 were present within the top 12 clusters of CR10. Albeit, the distribution and ranking differed between the two pools **(Table 6.3)**. Both pools had the same top cluster, C10-1. On the other hand, cluster 6 at CR10, which corresponds to C7-1, was ranked second in RS-C10. C10-2 and C10-3 were also within the top five cluster at RS-C10 **(Table 6.3)**. Interestingly, the CR10 clusters 10 and 11 were considerably more enriched in the reselection pool **(Table 6.3**). Overall, reselection with DnaK lead to a shift in the ranking of the top sequences.

Table 6.2 Summary of sequencing data for RS-C9 and RS-C10. Sequencing data from C-libraries after 2 and 3 rounds of reselection with DnaK, RS-C9 and RS-C10, respectively. The number of sequences read by the equipment for each library is represented as the total number of reads. The number of unique clusters per round was calculated by organizing the individual reads into unique clusters (families of similar sequences). Calculated average number of reads per cluster and the average frequency of clusters, and the frequency of the most abundant cluster in each round is presented.

Round	Total Number of Reads	Number of Unique Clusters	Average Number of Reads per Cluster	Average Frequency of Clusters (%)	Frequency of Top Cluster (%)
RS-C10	383728	199258	1.93	5.02×10^{-4}	17.24
RS-C9	506283	241718	2.09	4.14×10^{-4}	19.87

Table 6.3 Top 10 clusters at RS-C10. The top 10 clusters after three rounds of reselection with full-length DnaK (RS-C10) are presented, and their relative abundance is described as a percentage. The ranking to these clusters in round 10 of the original DnaK-CTD selection (CR10) is also presented.

Rank in RS-C10	Abundance	Sequence	Rank in CR10
1	17.244	CATCAAAGCTAGTCTTCCGAAAACGTGCCGTAAGGTGTGC	1
2	11.352	ACTGTCCCCTCTTTTTAATAGCCGCGTATAATGTTTGTCC	6
3	5.402	CCACCGCGAAAGACGGTCCCGAGCCCGCGAGTGGCACACA	3
4	4.289	GATACGGTCGCGTGAGACGACGTATGCCCAGCCGTTCACA	2
5	2.604	TGCGCGGGTGGATATCTGGCCAGGGAAGGGGGTGGTGGTG	10
6	1.961	TACACTTTGTCCGCCATAAGTGAATACGTTTGAGTACTAGC	4
7	1.360	AGGACGCGTGGTGGGATTTTCGGTGGATTTAGGTAGGGGG	11
8	0.590	TGGATTGGAGGATGTTTGCGCCGGTTCGAGGTCGCGTTTT	7
9	0.266	CATGGGTTTGGATATCGTCAACGGGAAGGGGGCTGGCGGG	12
10	0.194	TGTTTGATAGAAGATGTTTCCGCGTTCGGTGCCGGAGGGT	8

The enrichment patterns of the top 12 clusters at CR10 were examined to further compare the differences between the DnaK-CTD selection and the DnaK reselection. The frequencies of the clusters at CR9 were used as a baseline to assess the effect of the new target. Cluster one, C10-1, was found to be less abundant in the reselection pool **(Figure 6.3 A)**. While the frequency of clusters 2 to 5 (C10-2, C10-3, C10-4, and C10-5) did not change considerably **(Figure 6.3 A)**. In contrast, cluster 6 (C7-1) was substantially more abundant in the reselection libraries than in CR9 **(Figure 6.3 A)**. Cluster 7, 8 and 9, were also found in similar abundance in CR9 and the reselection pools **(Figure 6.3 B)**. Notably, clusters 10 to 12 enriched more significantly during the reselection **(Figure 6.3 B)**.

To further assess enrichment patterns the enrichment factor between RS-C9 and RS-C10 were calculated and plotted **(Figure 6.4 A)**. Clusters 1 to 5 and cluster 10 had a positive enrichment. Notably, the magnitude of the enrichment factors was on average smaller for positively enriched sequences (14.00) than that of the negatively enriched ones (-40.43). Enrichment factors were also calculated between CR9 and RS-C10 to compare the original selection to the reselection **(Figure 6.4 B)**. Clusters 1, 4, 5 and 9 were less abundant in the reselection pool, while the remaining clusters had positive enrichment factors.

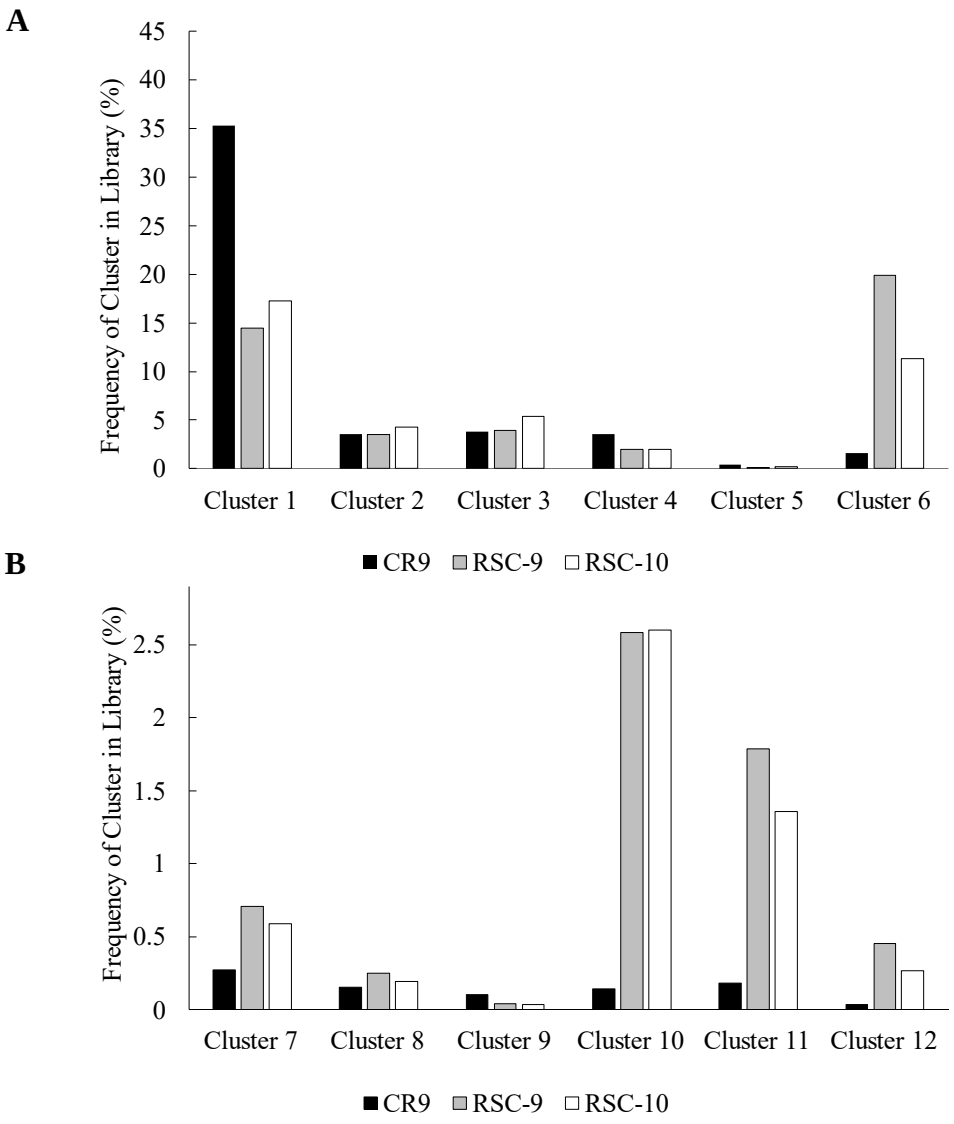


Figure 6.3 Effects of reselection with DnaK on the abundance of the top CR10 clusters. The top 12 clusters after selection with DnaK-CTD were identified and their frequency was tracked to the CR9 library and to the reselection libraries RS-C9 and RS-C10. The frequency of **(A)** clusters 1 to 6, and (**B)** clusters 7 to 12 are presented as percent abundance in library.

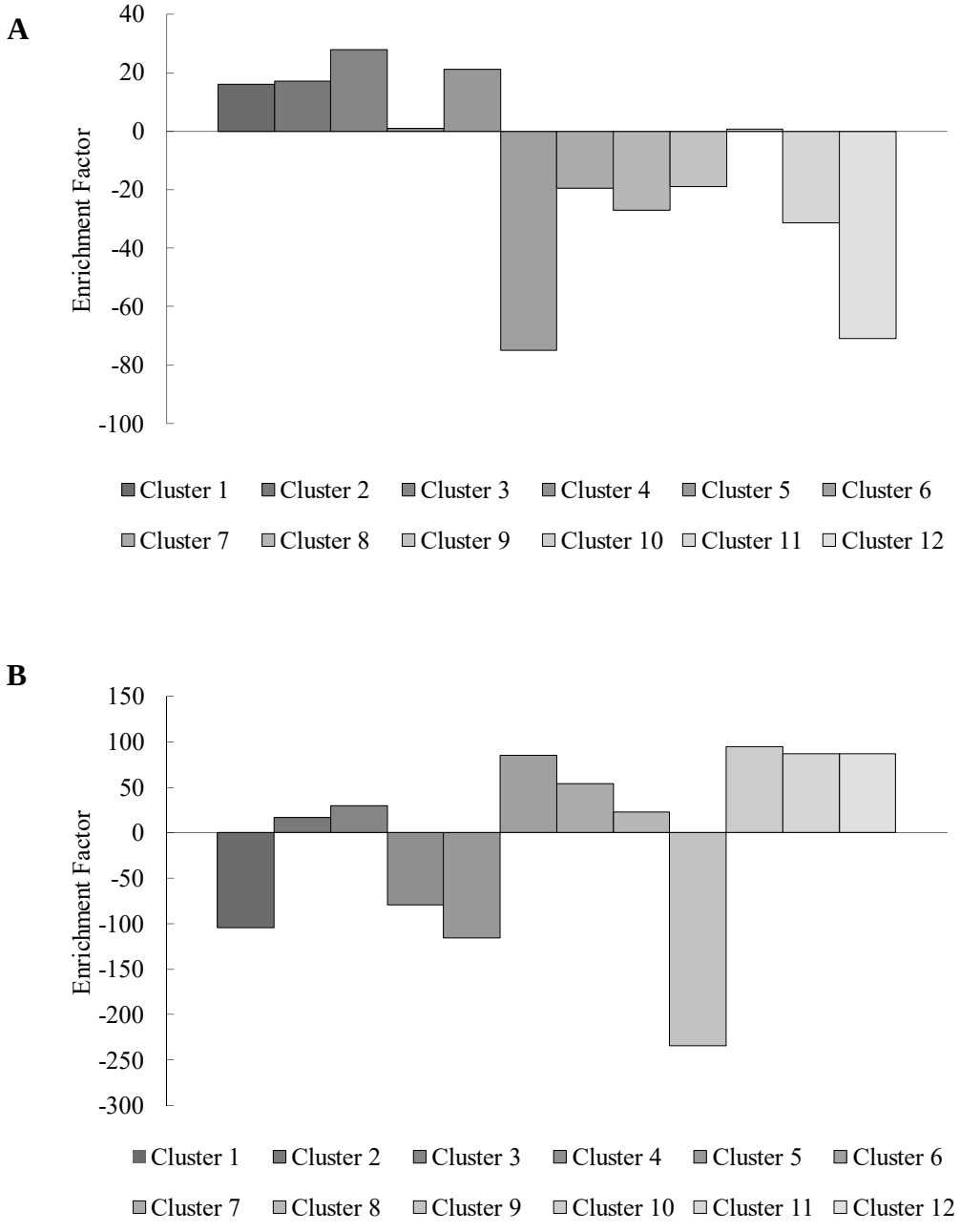


Figure 6.4 Enrichment of top 12 clusters. Enrichment factor of CR10 clusters 1 to 12 was calculated between **(A)** RS-C9 and RS-C10, and **(B)** CR9 and RS-C10. Clusters that are more abundant in RS-C10 have positive enrichment.

6.5. Discussion and conclusion

The aptamers isolated through the selections for DnaK-NTD (N-aptamers) and DnaK-CTD (C-aptamers) did not bind the full-length DnaK protein, even though they were able to recognize the individual domains **(Figure 6.1)**. Despite multiple attempts at troubleshooting the binding assays, these aptamers failed to recognize the full-length protein. Studies have demonstrated that DnaK has a highly dynamic structure and a conformational cycle; DnaK opens when bound to ATP, and upon ATP hydrolysis DnaK closes and has high affinity to bound peptides (Matthias P. Mayer & Gierasch, 2019). Therefore, it is possible that the DnaK conformations recognized by the aptamers are different to the main conformation adopted by the DnaK protein stock.

The functional cycle of DnaK is mediated by association with other chaperone proteins; association with DnaJ mediates substrate binding and ATP hydrolysis, and association with GrpE mediates nucleotide exchanging from ADP to ATP (Matthias P. Mayer & Gierasch, 2019). Thus, the DnaK stock protein is most likely in the ADP-bound conformation. Yet, it was previously reported that the conformation of ADP-bound DnaK allowed the domains to behave independently of each other (Bertelsen et al., 2009). Then, the conformation of the DnaK-NTD and DnaK-CTD should be like that of the DnaK in the stock. Nevertheless, the lack of aptamer binding towards DnaK suggests that the conformations of the individual domains and the full-length protein are not equivalent.

Another possibility is that DnaK may exist in an aggregated form in the protein stocks. DnaK has been reported to occur as monomers, dimers, trimers, and oligomers under *in vitro* and *in vivo* conditions (Sarbeng et al., 2015; Thompson et al., 2012). Oligomerization of DnaK has been associated to multiple factors. For instance, DnaK has been reported to exist mostly as a monomer in its ATP bound state, and as a mixture of oligomers in the ADP-bound state (Sarbeng et al., 2015; Thompson et al., 2012). As such, it is probable that the DnaK stocks contain a population of multimeric DnaK that cannot be recognized by the aptamers selected. However, structural studies would be necessary to reveal the reason why the aptamers selected are unable to recognize DnaK.

It is possible that the issue lies with the separation of the protein into individual domains and in the design of the constructs, and not on the DnaK stocks. To address this possibility, the reselection project was designed to conduct selection using DnaK for three rounds using the N-library at round 5, and the C-library at round 7. Reselection using the N-library did not show enrichment of sequences **(Figure 6.2)**. The original selection against DnaK-NTD enriched quickly, and at round 7 the top sequence made up 70 % of the library **(Table 4.4)**. The highly aptameric epitope of DnaK-NTD may not have been available for binding in DnaK. Then, the reselection project would need to continue for a few more rounds before enrichment occurs and DnaK binding aptamers arise. Thus, N7-1, N7-2 and N7-3 failed to bind DnaK in the dot-blot assay and in the reselection project.

The C-libraries enriched over reselection, albeit to a slightly lower degree than the original selection. Although no new sequences emerged from reselection, the ranking of top sequences and their enrichment patterns were different from the original selection. These data suggest that emergent aptamer sequences should be tested against DnaK, like clusters 6, 10 and 11. Yet, top sequences at RS-C10 (C10-1, C10-2 and C10-3) did not bind DnaK in the dot-blot assay, but were highly enriched in the reselection project. Therefore, the evidence regarding the C-aptamers is conflicting. Perhaps, further assessment of these sequences and the new aptamer candidates from the reselection project should still be considered as potential aptamers for DnaK.

Altogether, the evidence presented suggests that conducting selection in a domain-specific manner as presented in this thesis is not a good approach for DnaK. The lack of enrichment of the N-aptamers after reselection suggests that these sequences cannot recognize DnaK and that different aptamers will need to be identified to continue this project. While the conflicting evidence presented for the C-aptamers is puzzling. Diving into the literature for DnaK provides some insight on the structural complexity of this protein. In a simple model, DnaK conformation is dependent on binding partners, and nucleotide-binding status (Matthias P. Mayer & Gierasch, 2019). However, DnaK conformation can be influence by other factors. For instance, the post-translational modifications added by protein expression systems on DnaK have been shown to influence its structure

and function (Rigo et al., 2020). Hence, the structure of DnaK is highly sensitive to the conditions in which it is purified and how it is handled in solution.

Alternatively, the aptamer's recognition problems may be caused by discrepancies in the oligomerization state of the individual domains and the full protein. Oligomerization of DnaK has been proposed to be driven by the SBD domain located in the CTD (Aprile et al., 2013). It was reported that the truncated DnaK-CTD exists as a population of dimers, trimers, and tetramers, and that DnaK-NTD is more likely to be found as a monomer (Aprile et al., 2013). This may explain why the C-aptamers were more likely to bind DnaK in the reselection than the N-aptamers. Nonetheless, the structural complexity of DnaK has become a confounding factor in these data. The only conclusion that can be drawn based on these results is that the N- and C-aptamers recognize the individual domains of DnaK, and that the conformations of these domains are not compatible with the conformation of DnaK present in the stocks tested.

CHAPTER 7:CONCLUSION AND FUTURE DIRECTIONS

The main objective of this thesis was to test the capacity of the multidomain selection approach to efficiently generate aptamers against the domains of DnaK which could be used to create a bivalent aptamer. The aptamers isolated successfully recognized the individual domains of DnaK, but they did not recognize the full protein in the model tested. As discussed above, this may be due to incompatibility between the conformation of the domains of DnaK and that of DnaK in the protein stocks. However, whether the selected aptamers can recognize DnaK in a biologically relevant conformation is still not known.

Thus, future work in this project should include testing the aptamers selected against naturally occurring DnaK from *F. nucleatum* cultures. An assay able to estimate the affinity of aptamers against DnaK in culture can be used to compare the binding of the individual aptamers and that of a bivalent aptamer. However, designing and optimizing a bivalent aptamer would be difficult with this kind of set up as the net concentration of DnaK in solution is not known.

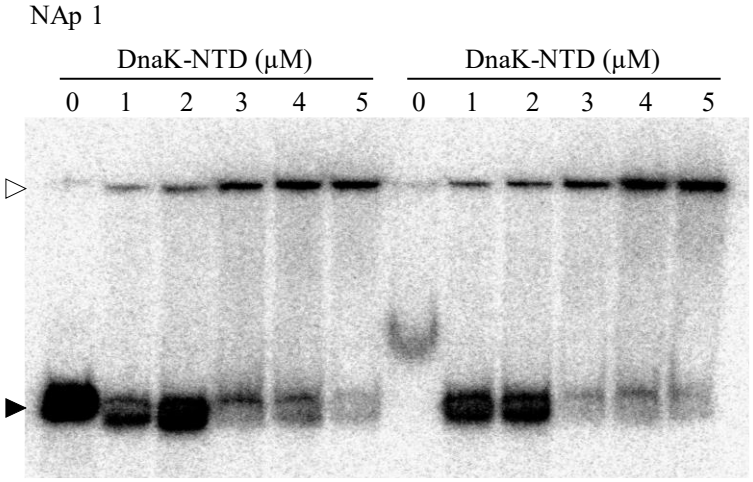
Multiple ongoing experiments were left unfinished due to the COVID-19 pandemic. For instance, the reselection project could be concluded by continuing reselection with the N-library for a few rounds of selection. Perhaps, new DnaK binding sequences would have emerged which could be characterized. Similarly, the C-library reselection project could be concluded by testing the emergent sequences against DnaK. After reselection, both the N- and the C-libraries should

be tested against DnaK and the individual domains to measure overall affinity of the DNA pools. If sequences that bind DnaK emerge, then, a bivalent aptamer could be built and characterized. However, it is likely that the issues with DnaK conformations discussed in the previous chapter would still be an issue.

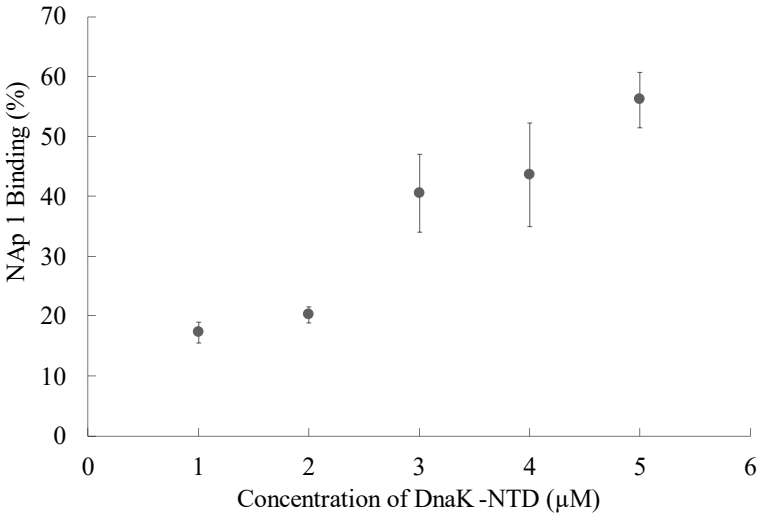
In summary, the evidence presented in this thesis is inconclusive on whether MDS can be used for creating high-affinity bivalent aptamers. Further work in the *in-vitro* model would be valuable in concluding the findings presented. Though, the *in-vitro* model of DnaK has been shown to be challenging. Future work should focus on characterizing the N- and C-aptamers against naturally occurring DnaK from *F. nucleatum* cultures.

SUPPLEMENTAL INFORMATION

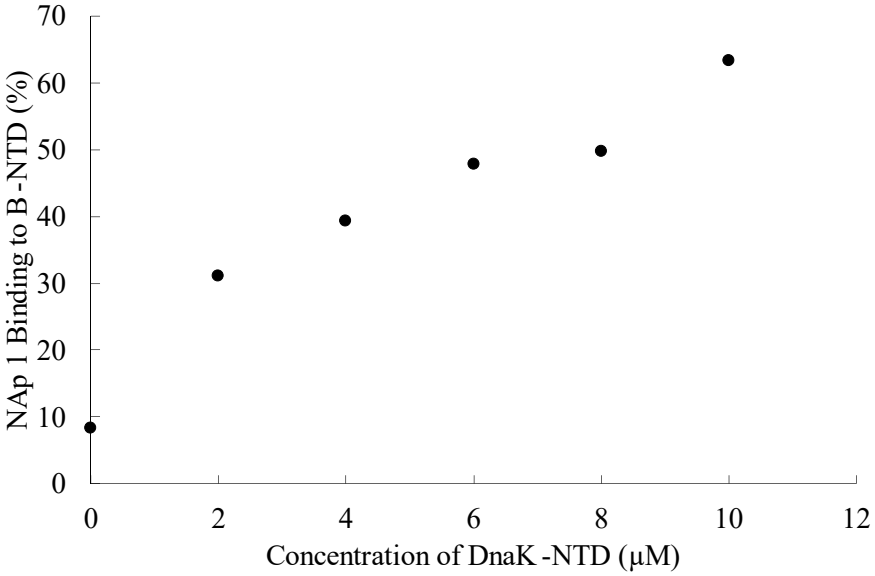
A.



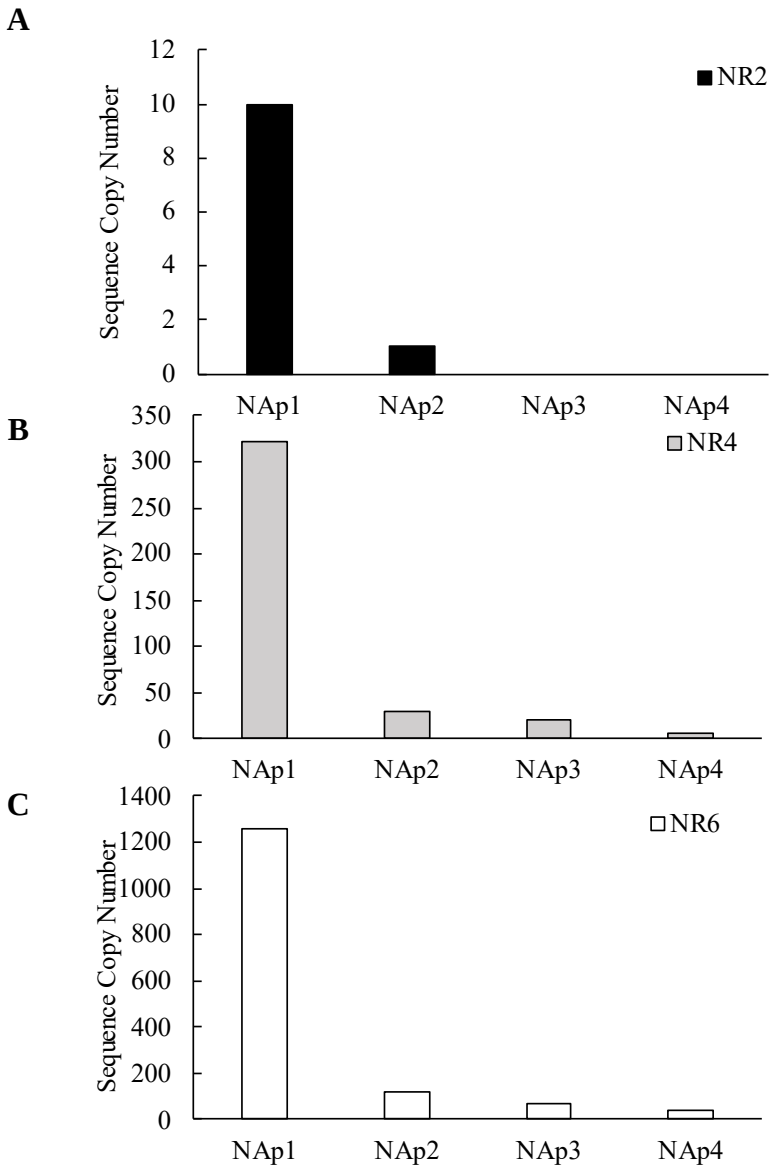
B.



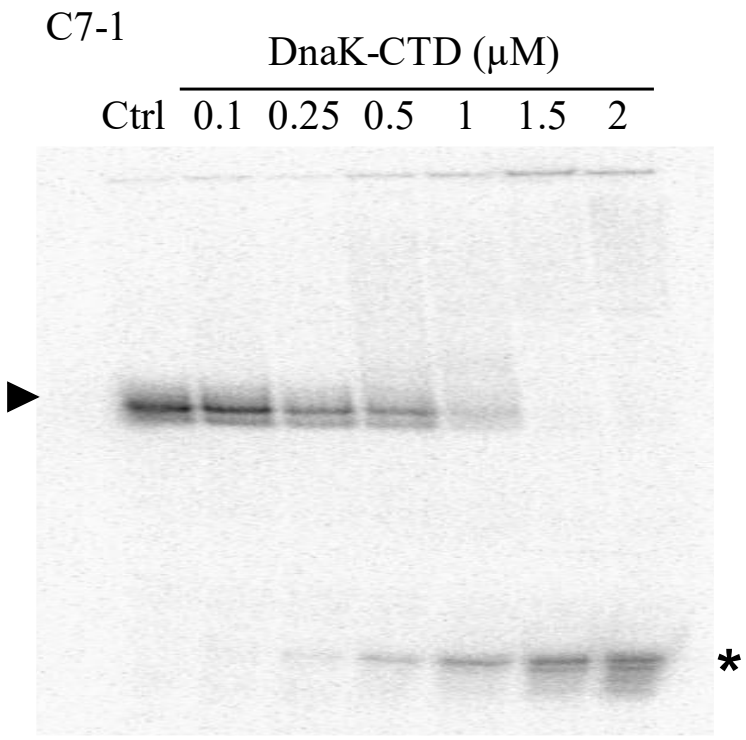
Supplemental Figure 1. Binding test of aptamers NAp 1 to *F. nucleatum* DnaK-NTD with electromobility shift assay (EMSA). (A) EMSA was performed with 25 nM of aptamer NAp 1 and various concentration of DnaK-NTD protein (0 μM to 5 μM). Aptamers were radiolabeled, incubated with target for one hour, and run on native PAGE (7 %). Gels were exposed to phosphor plates for 7 minutes and scanned with Typhoon. **(B)** Aptamer bound to the target was measured with ImageJ and a percent of aptamer binding was calculated. These data were used to create binding curves using GraphPad Prism non-linear regression for ligands with one-binding site. The dissociation constant (Kd) of NAp 1 was calculated to be 13.23 μM (R^2 0.9254).



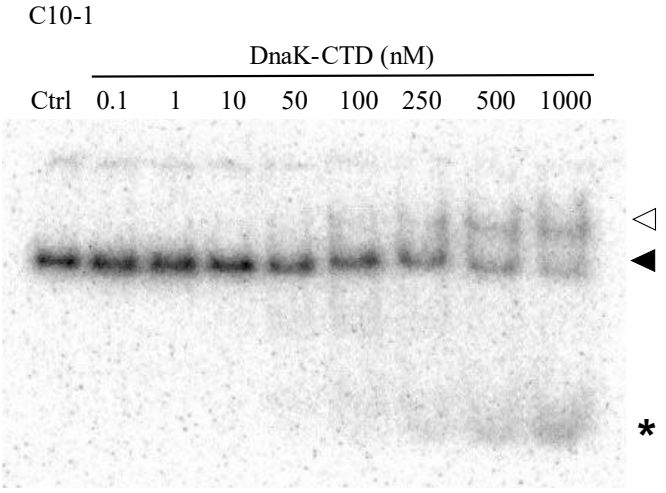
Supplemental Figure 2. Binding curve of aptamer NAp 1 to DnaK-NTD. Binding of aptamer NAp 1 to DnaK-NTD was tested with a bead binding test. DnaK-NTD was immobilized on beads (B-NTD) and diluted to various concentrations. Aptamer Nap1 was incubated with B-NTD for one hour. Bound and unbound aptamers were separated using a magnetic rack and run on a dPAGE. Binding of NAp1 to B-NTD was quantified with ImageJ and plotted. Binding data was analysed with GraphPad Prism and fitted to a non-linear fit for a one-site binding ligand. The Kd of NAp1 with the bead-binding assay was estimated to be 3.78 μM (R^2 0.8884).



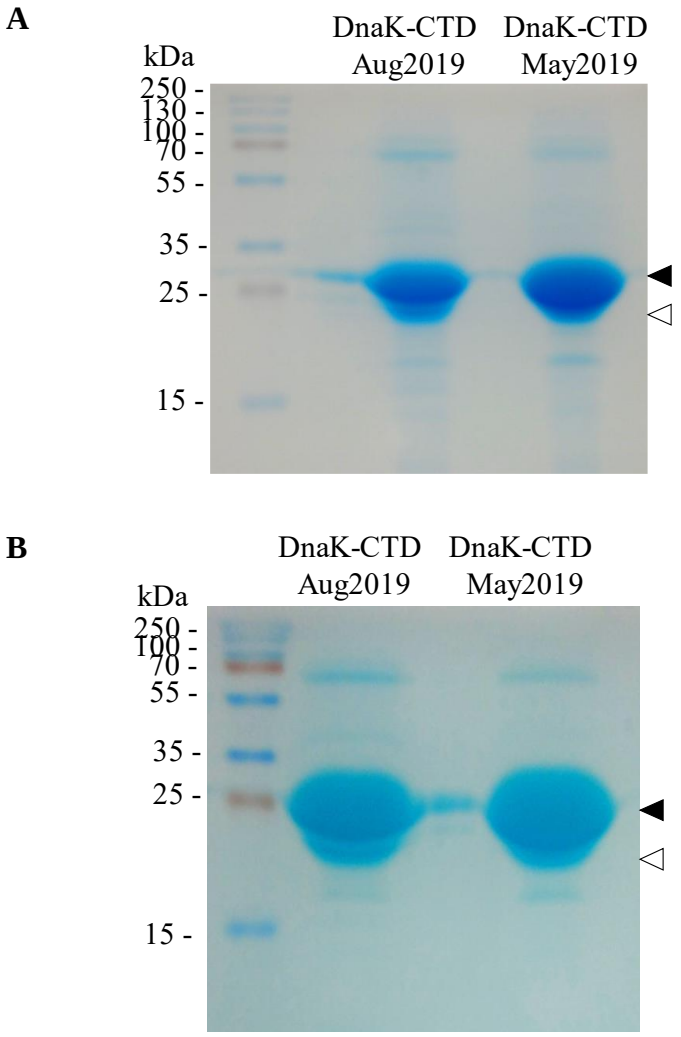
Supplemental Figure 3. Abundance of NAp aptamers in early rounds of selection. The abundance of the aptamer candidates in NAp 1, NAp 2, NAp 3 and NAp 4were tracked during the early rounds of selection **(A)** NR2, **(B)** NR4, and **(B)** NR6. Abundance of the sequences is measured as the number of reads in the DNA pool.



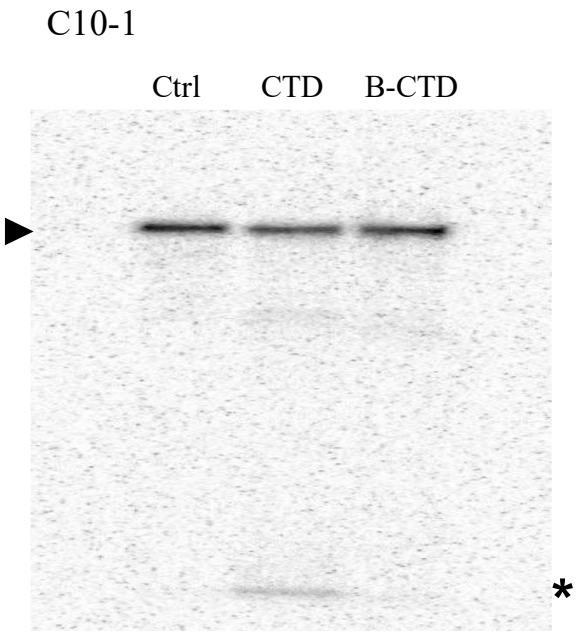
Supplemental Figure 4. Binding test of C7-1 to *F. nucleatum* DnaK-CTD with EMSA. Radiolabeled aptamer C7-1was incubated with various concentration of DnaK-CTD protein (0 μM to 2 μM) for one hour. The reaction buffer used did not contain Ca^{2+}. Samples were run on native PAGE (7 %). Gels were exposed to phosphor plates and scanned with Typhoon. Full-length aptamer is indicated by black arrowhead (►) and aptamer fragments are indicated by asterisk (*). Aptamer degradation was depended on the concentration of DnaK-CTD in solution.



Supplemental Figure 5. Binding test of C10-1 to *F. nucleatum* DnaK-CTD with EMSA. Radiolabeled aptamer C10-1was incubated with various concentration of DnaK-CTD protein (0 to 1000 nM) for one hour. The reaction buffer used did not contain Ca^{2+}. Samples were run on native PAGE (7 %). Gels were exposed to phosphor plates and scanned with Typhoon. Aptamer is indicated by the black arrowhead (◀) and aptamer bound to protein is indicated by the white arrowhead (◁). Aptamer fragments are noted by the asterisk (*). Aptamer degradation was dependent on the concentration of DnaK-CTD in solution**.**



Supplemental Figure 6. Purified *F. nucleatum* DnaK-CTD. Protein stocks of DnaK-CTD were purified using a HisTrap HP Ni-NTA affinity column, followed by an anion exchange column (HiTrap Q HP). Proteins were concentrated with centrifugal filters and 1 µL of concentrate was run on a **(A)** 15 % and **(B)** 20 % SDS PAGE. Gel was stained with Coomassie blue. Each lane represents a different purification experiment, DnaK-CTD from August 2019, or DnaK-CTD from May 2019. The protein band splits into two; the top protein is DnaK-CTD (◀) and the bottom protein is the lower molecular weight contaminant (◁).



Supplemental Figure 7. Nuclease activity is not present in B-CTD. Aptamer C10-1 was incubated with buffer (Ctrl), DnaK-CTD from stock (CTD) (1 µM), or DnaK-CTD immobilized on magnetic beads (B-CTD) (1 µM) for 1 hour. Aptamer in buffer was used as a negative control, and CTD was used as a positive control. Samples were run on a dPAGE (10%). Gel was exposed to storage phosphor and imaged with Typhoon. The position of the degraded aptamer is marked (◄).

REFERENCES

Ahmad, K. M., Xiao, Y., & Soh, H. T. (2012). Selection is more intelligent than design: Improving the affinity of a bivalent ligand through directed evolution. *Nucleic Acids Research*, *40*(22), 11777–11783. https://doi.org/10.1093/nar/gks899

Andrake, M., Guild, N., Hsu, T., Gold, L., Tuerk, C., & Karam, J. (1988). DNA polymerase of bacteriophage T4 is an autogenous translational repressor. *Proceedings of the National Academy of Sciences of the United States of America*, *85*(21), 7942–7946.

Aprile, F. A., Dhulesia, A., Stengel, F., Roodveldt, C., Benesch, J. L. P., Tortora, P., Robinson, C. V., Salvatella, X., Dobson, C. M., & Cremades, N. (2013). Hsp70 Oligomerization Is Mediated by an Interaction between the Interdomain Linker and the Substrate-Binding Domain. *PLoS ONE*, *8*(6), e67961. https://doi.org/10.1371/journal.pone.0067961

Arrabi, A. (2017). *A Multipronged Approach in Targeting Clostridium difficile: Multiple Domain Selection for Aptamer Isolation* [Msc]. McMaster University.

Bazak, R., Houri, M., El Achy, S., Kamel, S., & Refaat, T. (2015). Cancer active targeting by nanoparticles: A comprehensive review of literature. *Journal of Cancer Research and Clinical Oncology*, *141*(5), 769–784. https://doi.org/10.1007/s00432-014-1767-3

Bertelsen, E. B., Chang, L., Gestwicki, J. E., & Zuiderweg, E. R. P. (2009). Solution conformation of wild-type E. coli Hsp70 (DnaK) chaperone complexed with ADP and substrate. *Proceedings of the National Academy of Sciences*, *106*(21), 8471–8476. https://doi.org/10.1073/pnas.0903503106

Bock, L. C., Griffin, L. C., Latham, J. A., Vermaas, E. H., & Toole, J. J. (1992). Selection of single-stranded DNA molecules that bind and inhibit human thrombin. *Nature*, *355*(6360), 564–566. https://doi.org/10.1038/355564a0

Bolstad, A. I., Jensen, H. B., & Bakken, V. (1996). Taxonomy, biology, and periodontal aspects of Fusobacterium nucleatum. *Clinical Microbiology Reviews*, *9*(1), 55–71. https://doi.org/10.1128/CMR.9.1.55

Breaker, R. R., & Joyce, G. F. (1994). A DNA enzyme that cleaves RNA. *Chemistry & Biology*, *1*(4), 223–229. https://doi.org/10.1016/1074-5521(94)90014-0

Brennan, C. A., & Garrett, W. S. (2019). Fusobacterium nucleatum—Symbiont, opportunist and oncobacterium. *Nature Reviews Microbiology*, *17*(3), 156–166. https://doi.org/10.1038/s41579-018-0129-6

Bruno, J. G. (1997). In VitroSelection of DNA to Chloroaromatics Using Magnetic Microbead-Based Affinity Separation and Fluorescence Detection. *Biochemical and Biophysical Research Communications*, *234*(1), 117–120. https://doi.org/10.1006/bbrc.1997.6517

Buchberger, A., Theyssen, H., Schröder, H., McCarty, J. S., Virgallita, G., Milkereit, P., Reinstein, J., & Bukau, B. (1995). Nucleotide-induced Conformational Changes in the ATPase and Substrate Binding Domains of the DnaK Chaperone Provide Evidence for Interdomain Communication. *Journal of Biological Chemistry*, *270*(28), 16903–16910. https://doi.org/10.1074/jbc.270.28.16903

Castellarin, M., Warren, R. L., Freeman, J. D., Dreolini, L., Krzywinski, M., Strauss, J., Barnes, R., Watson, P., Allen-Vercoe, E., Moore, R. A., & Holt, R. A. (2012). Fusobacterium nucleatum infection is prevalent in human colorectal carcinoma. *Genome Research*, *22*(2), 299–306. https://doi.org/10.1101/gr.126516.111

Chen, Y.-C., Li, C.-L., Hsiao, Y.-Y., Duh, Y., & Yuan, H. S. (2014). Structure and function of TatD exonuclease in DNA repair. *Nucleic Acids Research*, *42*(16), 10776–10785. https://doi.org/10.1093/nar/gku732

Dai, Z., Coker, O. O., Nakatsu, G., Wu, W. K. K., Zhao, L., Chen, Z., Chan, F. K. L., Kristiansen, K., Sung, J. J. Y., Wong, S. H., & Yu, J. (2018). Multi-cohort analysis of colorectal cancer metagenome identified altered bacteria across populations and universal bacterial markers. *Microbiome*, *6*(1), 70. https://doi.org/10.1186/s40168-018-0451-2

de Carvalho, A. C., de Mattos Pereira, L., Datorre, J. G., dos Santos, W., Berardinelli, G. N., Matsushita, M. de M., Oliveira, M. A., Durães, R. O., Guimarães, D. P., & Reis, R. M. (2019). Microbiota Profile and Impact of Fusobacterium nucleatum in Colorectal Cancer Patients of Barretos Cancer Hospital. *Frontiers in Oncology*, *9*, 813. https://doi.org/10.3389/fonc.2019.00813

Dhiman, A., Kalra, P., Bansal, V., Bruno, John. G., & Sharma, T. K. (2017). Aptamer-based point-of-care diagnostic platforms. *Sensors and Actuators B: Chemical*, *246*, 535–553. https://doi.org/10.1016/j.snb.2017.02.060

Dong, Y., Zhang, T., Lin, X., Feng, J., Luo, F., Gao, H., Wu, Y., Deng, R., & He, Q. (2020). Graphene/aptamer probes for small molecule detection: From in vitro test to in situ imaging. *Microchimica Acta*, *187*(3), 179. https://doi.org/10.1007/s00604-020-4128-8

Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792–1797. https://doi.org/10.1093/nar/gkh340

Edgar, Robert C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, *26*(19), 2460–2461. https://doi.org/10.1093/bioinformatics/btq461

Ellington, A. D., & Szostak, J. W. (1990). In vitro selection of RNA molecules that bind specific ligands. *Nature*, *346*(6287), 818–822. https://doi.org/10.1038/346818a0

Fallah, A., Sadeghinia, A., Kahroba, H., Samadi, A., Heidari, H. R., Bradaran, B., Zeinali, S., & Molavi, O. (2019). Therapeutic targeting of angiogenesis molecular pathways in angiogenesis-dependent diseases. *Biomedicine &*

Pharmacotherapy, *110*, 775–785. https://doi.org/10.1016/j.biopha.2018.12.022

Flaherty, K. M., DeLuca-Flaherty, C., & McKay, D. B. (1990). Three-dimensional structure of the ATPase fragment of a 70K heat-shock cognate protein. *Nature*, *346*(6285), 623–628. https://doi.org/10.1038/346623a0

Flanagan, L., Schmid, J., Ebert, M., Soucek, P., Kunicka, T., Liska, V., Bruha, J., Neary, P., Dezeeuw, N., Tommasino, M., Jenab, M., Prehn, J. H. M., & Hughes, D. J. (2014). Fusobacterium nucleatum associates with stages of colorectal neoplasia development, colorectal cancer and disease outcome. *European Journal of Clinical Microbiology & Infectious Diseases*, *33*(8), 1381–1390. https://doi.org/10.1007/s10096-014-2081-3

Flynn-Charlebois, A., Wang, Y., Prior, T. K., Rashid, I., Hoadley, K. A., Coppins, R. L., Wolf, A. C., & Silverman, S. K. (2003). Deoxyribozymes with 2‘−5‘ RNA Ligase Activity. *Journal of the American Chemical Society*, *125*(9), 2444–2454. https://doi.org/10.1021/ja028774y

Gaiser, R. A., Halimi, A., Alkharaan, H., Lu, L., Davanian, H., Healy, K., Hugerth, L. W., Ateeb, Z., Valente, R., Fernández Moro, C., Del Chiaro, M., & Sällberg Chen, M. (2019). Enrichment of oral microbiota in early cystic precursors to invasive pancreatic cancer. *Gut*, *68*(12), 2186–2194. https://doi.org/10.1136/gutjnl-2018-317458

Gao, R., Zhong, Z., Gao, X., & Jia, L. (2018). Graphene Oxide Quantum Dots Assisted Construction of Fluorescent Aptasensor for Rapid Detection of *Pseudomonas aeruginosa* in Food Samples. *Journal of Agricultural and Food Chemistry*, *66*(41), 10898–10905. https://doi.org/10.1021/acs.jafc.8b02164

Gauss, P., Gayle, M., Winter, R. B., & Gold, L. (1987). The bacteriophage T4 dexA gene: Sequence and analysis of a gene conditionally required for DNA replication. *MGG Molecular & General Genetics*, *206*(1), 24–34. https://doi.org/10.1007/BF00326532

Gelinas, A. D., Davies, D. R., & Janjic, N. (2016). Embracing proteins: Structural themes in aptamer–protein complexes. *Current Opinion in Structural Biology*, *36*, 122–132. https://doi.org/10.1016/j.sbi.2016.01.009

Gholizadeh, P., Eslami, H., Yousefi, M., Asgharzadeh, M., Aghazadeh, M., & Kafil, H. S. (2016). Role of oral microbiome on oral cancers, a review. *Biomedicine & Pharmacotherapy*, *84*, 552–558. https://doi.org/10.1016/j.biopha.2016.09.082

Gong, Q., Wang, J., Ahmad, K. M., Csordas, A. T., Zhou, J., Nie, J., Stewart, R., Thomson, J. A., Rossi, J. J., & Soh, H. T. (2012). Selection Strategy to Generate Aptamer Pairs that Bind to Distinct Sites on Protein Targets. *Analytical Chemistry*, *84*(12), 5365–5371. https://doi.org/10.1021/ac300873p

Grobbee, E. J., Lam, S. Y., Fuhler, G. M., Blakaj, B., Konstantinov, S. R., Bruno, M. J., Peppelenbosch, M. P., Kuipers, E. J., & Spaander, M. C. (2019).

First steps towards combining faecal immunochemical testing with the gut microbiome in colorectal cancer screening. *United European Gastroenterology Journal*, 205064061989073. https://doi.org/10.1177/2050640619890732

Guo, S., Li, L., Xu, B., Li, M., Zeng, Q., Xiao, H., Xue, Y., Wu, Y., Wang, Y., Liu, W., & Zhang, G. (2018). A Simple and Novel Fecal Biomarker for Colorectal Cancer: Ratio of Fusobacterium Nucleatum to Probiotics Populations, Based on Their Antagonistic Effect. *Clinical Chemistry*, *64*(9), 1327–1337. https://doi.org/10.1373/clinchem.2018.289728

Hasegawa, H., Taira, K., Sode, K., & Ikebukuro, K. (2008). Improvement of Aptamer Affinity by Dimerization. *Sensors*, *8*(12), 1090–1098. https://doi.org/10.3390/s8021090

Kim, Y., Cao, Z., & Tan, W. (2008). Molecular assembly for high-performance bivalent nucleic acid inhibitor. *Proceedings of the National Academy of Sciences*, *105*(15), 5664–5669. https://doi.org/10.1073/pnas.0711803105

Kityk, R., Kopp, J., Sinning, I., & Mayer, M. P. (2012). Structure and Dynamics of the ATP-Bound Open Conformation of Hsp70 Chaperones. *Molecular Cell*, *48*(6), 863–874. https://doi.org/10.1016/j.molcel.2012.09.023

Kostic, A. D., Gevers, D., Pedamallu, C. S., Michaud, M., Duke, F., Earl, A. M., Ojesina, A. I., Jung, J., Bass, A. J., Tabernero, J., Baselga, J., Liu, C., Shivdasani, R. A., Ogino, S., Birren, B. W., Huttenhower, C., Garrett, W. S., & Meyerson, M. (2012). Genomic analysis identifies association of Fusobacterium with colorectal carcinoma. *Genome Research*, *22*(2), 292–298. https://doi.org/10.1101/gr.126573.111

Kruger, K., Grabowski, P. J., Zaug, A. J., Sands, J., Gottschling, D. E., & Cech, T. R. (1982). Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of tetrahymena. *Cell*, *31*(1), 147–157. https://doi.org/10.1016/0092-8674(82)90414-7

Kyrgiou, M., Mitra, A., & Moscicki, A.-B. (2017). Does the vaginal microbiota play a role in the development of cervical cancer? *Translational Research*, *179*, 168–182. https://doi.org/10.1016/j.trsl.2016.07.004

Lauka, L., Reitano, E., Carra, M. C., Gaiani, F., Gavriilidis, P., Brunetti, F., de'Angelis, G. L., Sobhani, I., & de'Angelis, N. (2019). Role of the intestinal microbiome in colorectal cancer surgery outcomes. *World Journal of Surgical Oncology*, *17*(1), 204. https://doi.org/10.1186/s12957-019-1754-x

Lee, D.-W., Han, S.-W., Kang, J.-K., Bae, J. M., Kim, H.-P., Won, J.-K., Jeong, S.-Y., Park, K. J., Kang, G. H., & Kim, T.-Y. (2018). Association Between Fusobacterium nucleatum, Pathway Mutation, and Patient Prognosis in Colorectal Cancer. *Annals of Surgical Oncology*, *25*(11), 3389–3395. https://doi.org/10.1245/s10434-018-6681-5

Lee, J. F., Hesselberth, J. R., Meyers, L. A., & Ellington, A. D. (2004). Aptamer database. *Nucleic Acids Research*, *32*(Database issue), D95-100. https://doi.org/10.1093/nar/gkh094

Li, Y., Liu, Y., & Breaker, R. R. (2000). Capping DNA with DNA. *Biochemistry*, *39*(11), 3106–3114.

Liang, Q., Chiu, J., Chen, Y., Huang, Y., Higashimori, A., Fang, J., Brim, H., Ashktorab, H., Ng, S. C., Ng, S. S. M., Zheng, S., Chan, F. K. L., Sung, J. J. Y., & Yu, J. (2017). Fecal Bacteria Act as Novel Biomarkers for Noninvasive Diagnosis of Colorectal Cancer. *Clinical Cancer Research*, *23*(8), 2061–2070. https://doi.org/10.1158/1078-0432.CCR-16-1599

Liang, Y. M., Wei, R. X., Hsu, T., Alford, C., Dawson, M., & Karam, J. (1988). Autogenous regulation of the regA gene of bacteriophage T4: Derepression of translation. *Genetics*, *119*(4), 743–749.

Liu, W., Zhang, R., Shu, R., Yu, J., Li, H., Long, H., Jin, S., Li, S., Hu, Q., Yao, F., Zhou, C., Huang, Q., Hu, X., Chen, M., Hu, W., Wang, Q., Fang, S., & Wu, Q. (2020). Study of the Relationship between Microbiome and Colorectal Cancer Susceptibility Using 16SrRNA Sequencing. *BioMed Research International*, *2020*, 1–17. https://doi.org/10.1155/2020/7828392

Mammen, M., Choi, S.-K., & Whitesides, G. M. (1998). Polyvalent Interactions in Biological Systems: Implications for Design and Use of Multivalent Ligands and Inhibitors. *Angewandte Chemie International Edition*, *37*(20), 2754–2794. https://doi.org/10.1002/(SICI)1521-3773(19981102)37:20<2754::AID-ANIE2754>3.0.CO;2-3

Manochehry, S., McConnell, E. M., & Li, Y. (2019). Unraveling Determinants of Affinity Enhancement in Dimeric Aptamers for a Dimeric Protein. *Scientific Reports*, *9*(1), 17824. https://doi.org/10.1038/s41598-019-54005-4

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, *17*(1), 10. https://doi.org/10.14806/ej.17.1.200

Mayer, G. (2009). The Chemical Biology of Aptamers. *Angewandte Chemie International Edition*, *48*(15), 2672–2689. https://doi.org/10.1002/anie.200804643

Mayer, M. P., & Bukau, B. (2005). Hsp70 chaperones: Cellular functions and molecular mechanism. *Cellular and Molecular Life Sciences*, *62*(6), 670–684. https://doi.org/10.1007/s00018-004-4464-6

Mayer, Matthias P. (2013). Hsp70 chaperone dynamics and molecular mechanism. *Trends in Biochemical Sciences*, *38*(10), 507–514. https://doi.org/10.1016/j.tibs.2013.08.001

Mayer, Matthias P., & Gierasch, L. M. (2019). Recent advances in the structural and mechanistic aspects of Hsp70 molecular chaperones. *Journal of Biological Chemistry*, *294*(6), 2085–2097. https://doi.org/10.1074/jbc.REV118.002810

McConnell, E. M., Cozma, I., Morrison, D., & Li, Y. (2020). Biosensors Made of Synthetic Functional Nucleic Acids Toward Better Human Health. *Analytical Chemistry*, *92*(1), 327–344. https://doi.org/10.1021/acs.analchem.9b04868

McConnell, E. M., Morrison, D., Rey Rincon, M. A., Salena, B. J., & Li, Y. (2020). Selection and applications of synthetic functional DNAs for bacterial detection. *TrAC Trends in Analytical Chemistry*, *124*, 115785. https://doi.org/10.1016/j.trac.2019.115785

McCoy, A. N., Araújo-Pérez, F., Azcárate-Peril, A., Yeh, J. J., Sandler, R. S., & Keku, T. O. (2013). Fusobacterium Is Associated with Colorectal Adenomas. *PLoS ONE*, *8*(1), e53653. https://doi.org/10.1371/journal.pone.0053653

Mima, K., Nishihara, R., Qian, Z. R., Cao, Y., Sukawa, Y., Nowak, J. A., Yang, J., Dou, R., Masugi, Y., Song, M., Kostic, A. D., Giannakis, M., Bullman, S., Milner, D. A., Baba, H., Giovannucci, E. L., Garraway, L. A., Freeman, G. J., Dranoff, G., … Ogino, S. (2016). *Fusobacterium nucleatum* in colorectal carcinoma tissue and patient prognosis. *Gut*, *65*(12), 1973–1980. https://doi.org/10.1136/gutjnl-2015-310101

Mizutani, S., Yamada, T., & Yachida, S. (2020). Significance of the gut microbiome in multistep colorectal carcinogenesis. *Cancer Science*, *111*(3), 766–773. https://doi.org/10.1111/cas.14298

Müller, J., Freitag, D., Mayer, G., & PöTzsch, B. (2008). Anticoagulant characteristics of HD1-22, a bivalent aptamer that specifically inhibits thrombin and prothrombinase. *Journal of Thrombosis and Haemostasis*, *6*(12), 2105–2112. https://doi.org/10.1111/j.1538-7836.2008.03162.x

Müller, Jens, Wulffen, B., Pötzsch, B., & Mayer, G. (2007). Multidomain Targeting Generates a High-Affinity Thrombin-Inhibiting Bivalent Aptamer. *ChemBioChem*, *8*(18), 2223–2226. https://doi.org/10.1002/cbic.200700535

Nonaka, Y., Sode, K., & Ikebukuro, K. (2010). Screening and Improvement of an Anti-VEGF DNA Aptamer. *Molecules*, *15*(12), 215–225. https://doi.org/10.3390/molecules15010215

Reid, R., Chatterjee, B., Das, S. J., Ghosh, S., & Sharma, T. K. (2020). Application of aptamers as molecular recognition elements in lateral flow assays. *Analytical Biochemistry*, *593*, 113574. https://doi.org/10.1016/j.ab.2020.113574

Robertson, D. L., & Joyce, G. F. (1990). Selection in vitro of an RNA enzyme that specifically cleaves single-stranded DNA. *Nature*, *344*(6265), 467–468. https://doi.org/10.1038/344467a0

Russo, E., Bacci, G., Chiellini, C., Fagorzi, C., Niccolai, E., Taddei, A., Ricci, F., Ringressi, M. N., Borrelli, R., Melli, F., Miloeva, M., Bechi, P., Mengoni, A., Fani, R., & Amedei, A. (2018). Preliminary Comparison of Oral and Intestinal Human Microbiota in Patients with Colorectal Cancer: A Pilot Study. *Frontiers in Microbiology*, *8*, 2699. https://doi.org/10.3389/fmicb.2017.02699

Sedighian, H., Halabian, R., Amani, J., Heiat, M., Amin, M., & Fooladi, A. A. I. (2018). Staggered Target SELEX, a novel approach to isolate non-cross-

reactive aptamer for detection of SEA by apta-qPCR. *Journal of Biotechnology*, *286*, 45–55. https://doi.org/10.1016/j.jbiotec.2018.09.006

Srinivasarao, M., & Low, P. S. (2017). Ligand-Targeted Drug Delivery. *Chemical Reviews*, *117*(19), 12133–12164. https://doi.org/10.1021/acs.chemrev.7b00013

Tian, L., & Heyduk, T. (2009). Bivalent Ligands with Long Nanometer-Scale Flexible Linkers †. *Biochemistry*, *48*(2), 264–275. https://doi.org/10.1021/bi801630b

Tucker, W., Kinghorn, A., Fraser, L., Cheung, Y.-W., & Tanner, J. (2018). Selection and Characterization of a DNA Aptamer Specifically Targeting Human HECT Ubiquitin Ligase WWP1. *International Journal of Molecular Sciences*, *19*(3), 763. https://doi.org/10.3390/ijms19030763

Tuerk, C., & Gold, L. (1990). Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science (New York, N.Y.)*, *249*(4968), 505–510.

Tunsjø, H. S., Gundersen, G., Rangnes, F., Noone, J. C., Endres, A., & Bemanian, V. (2019). Detection of Fusobacterium nucleatum in stool and colonic tissues from Norwegian colorectal cancer patients. *European Journal of Clinical Microbiology & Infectious Diseases*, *38*(7), 1367–1376. https://doi.org/10.1007/s10096-019-03562-7

Vauquelin, G., & Charlton, S. J. (2013). Exploring avidity: Understanding the potential gains in functional affinity and target residence time of bivalent and heterobivalent ligands: Exploring bivalent ligand binding properties. *British Journal of Pharmacology*, *168*(8), 1771–1785. https://doi.org/10.1111/bph.12106

Vogtmann, E., Hua, X., Zeller, G., Sunagawa, S., Voigt, A. Y., Hercog, R., Goedert, J. J., Shi, J., Bork, P., & Sinha, R. (2016). Colorectal Cancer and the Human Gut Microbiome: Reproducibility with Whole-Genome Shotgun Sequencing. *PLOS ONE*, *11*(5), e0155362. https://doi.org/10.1371/journal.pone.0155362

Wang, Y., & Silverman, S. K. (2003). Deoxyribozymes That Synthesize Branched and Lariat RNA. *Journal of the American Chemical Society*, *125*(23), 6880–6881. https://doi.org/10.1021/ja035150z

Wei, Z., Cao, S., Liu, S., Yao, Z., Sun, T., Li, Y., Li, J., Zhang, D., & Zhou, Y. (2016). Could gut microbiota serve as prognostic biomarker associated with colorectal cancer patients' survival? A pilot study on relevant mechanism. *Oncotarget*, *7*(29). https://doi.org/10.18632/oncotarget.10064

Wong, S. H., Kwong, T. N. Y., Chow, T.-C., Luk, A. K. C., Dai, R. Z. W., Nakatsu, G., Lam, T. Y. T., Zhang, L., Wu, J. C. Y., Chan, F. K. L., Ng, S. S. M., Wong, M. C. S., Ng, S. C., Wu, W. K. K., Yu, J., & Sung, J. J. Y. (2017). Quantitation of faecal Fusobacterium improves faecal immunochemical test in detecting advanced colorectal neoplasia. *Gut*, *66*(8), 1441–1448. https://doi.org/10.1136/gutjnl-2016-312766

Yachida, S., Mizutani, S., Shiroma, H., Shiba, S., Nakajima, T., Sakamoto, T., Watanabe, H., Masuda, K., Nishimoto, Y., Kubo, M., Hosoda, F., Rokutan, H., Matsumoto, M., Takamaru, H., Yamada, M., Matsuda, T., Iwasaki, M., Yamaji, T., Yachida, T., … Yamada, T. (2019). Metagenomic and metabolomic analyses reveal distinct stage-specific phenotypes of the gut microbiota in colorectal cancer. *Nature Medicine*, *25*(6), 968–976. https://doi.org/10.1038/s41591-019-0458-7

Yang, Y., Cai, Q., Shu, X., Steinwandel, M. D., Blot, W. J., Zheng, W., & Long, J. (2019). Prospective study of oral microbiome and colorectal cancer risk in low-income and African American populations. *International Journal of Cancer*, *144*(10), 2381–2389. https://doi.org/10.1002/ijc.31941

Yu, J., Feng, Q., Wong, S. H., Zhang, D., Liang, Q. yi, Qin, Y., Tang, L., Zhao, H., Stenvang, J., Li, Y., Wang, X., Xu, X., Chen, N., Wu, W. K. K., Al-Aama, J., Nielsen, H. J., Kiilerich, P., Jensen, B. A. H., Yau, T. O., … Wang, J. (2017). Metagenomic analysis of faecal microbiome as a tool towards targeted non-invasive biomarkers for colorectal cancer. *Gut*, *66*(1), 70–78. https://doi.org/10.1136/gutjnl-2015-309800

Zackular, J. P., Rogers, M. A. M., Ruffin, M. T., & Schloss, P. D. (2014). The Human Gut Microbiome as a Screening Tool for Colorectal Cancer. *Cancer Prevention Research*, *7*(11), 1112–1121. https://doi.org/10.1158/1940-6207.CAPR-14-0129

Zeller, G., Tap, J., Voigt, A. Y., Sunagawa, S., Kultima, J. R., Costea, P. I., Amiot, A., Böhm, J., Brunetti, F., Habermann, N., Hercog, R., Koch, M., Luciani, A., Mende, D. R., Schneider, M. A., Schrotz-King, P., Tournigand, C., Tran Van Nhieu, J., Yamada, T., … Bork, P. (2014). Potential of fecal microbiota for early-stage detection of colorectal cancer. *Molecular Systems Biology*, *10*(11), 766. https://doi.org/10.15252/msb.20145645

Zhang, X., Zhu, X., Cao, Y., Fang, J.-Y., Hong, J., & Chen, H. (2019). Fecal *Fusobacterium nucleatum* for the diagnosis of colorectal tumor: A systematic review and meta-analysis. *Cancer Medicine*, *8*(2), 480–491. https://doi.org/10.1002/cam4.1850

Zhao, X., Lis, J. T., & Shi, H. (2013). A Systematic Study of the Features Critical for Designing a High Avidity Multivalent Aptamer. *Nucleic Acid Therapeutics*, *23*(3), 238–242. https://doi.org/10.1089/nat.2012.0410

Zhu, X., Zhao, X., Burkholder, W. F., Gragerov, A., Ogata, C. M., Gottesman, M. E., & Hendrickson, W. A. (1996). Structural analysis of substrate binding by the molecular chaperone DnaK. *Science (New York, N.Y.)*, *272*(5268), 1606–1614.