Characterization of DNA binding of the two zinc finger domains of transcription factor zBED6

Alexander Taubert

Degree project in biology, Master of science (2 years), 2019
Examensarbete i biologi 30 hp till masterexamen, 2019
Biology Education Centre and Cell and Molecular Biology, Uppsala University
Supervisors: Maria Selmer and Sandesh Kanchugal Puttaswamy
External opponent: Francisco Javier Marcos-Torres
Table of Contents
Abstract........................................................................................................................................... 2
List of Abbreviations ...................................................................................................................... 3
1. Introduction.................................................................................................................................. 4
  1.1 The zBED6 transcription factor .......................................................................................... 4
  1.2 Zinc Finger Structure and Function .................................................................................. 6
1.3 Protein Purification ....................................................................................................................... 7
  1.3.1 Immobilized metal ion affinity chromatography .............................................................. 7
  1.3.2 Size exclusion chromatography ...................................................................................... 7
1.4 Methods to characterize DNA-protein interactions ................................................................. 8
  1.4.1 Electrophoretic Mobility Shift Assay ............................................................................ 8
  1.4.2 Microscale Thermophoresis ......................................................................................... 8
2. Aims ............................................................................................................................................ 9
3. Methods ..................................................................................................................................... 10
  3.1 Transformation of plasmid .................................................................................................... 10
  3.1.1 Transformation to BL21(AI) and BL21(DE3) pLysS expression cells ......................... 10
  3.1.2 Transformation to TOP10 competent cells and mini-prep ............................................ 10
  3.2 Protein Purification .............................................................................................................. 11
    3.2.1 Small Scale Expression Test ...................................................................................... 11
    3.2.2 Large Scale Expression .............................................................................................. 11
    3.2.3 Protein Purification .................................................................................................... 12
    3.2.4 SDS-PAGE Protocol ................................................................................................. 13
  3.3 Electrophoretic Mobility Shift Assay .................................................................................... 13
    3.3.1 Annealing of Double-Stranded DNA ........................................................................ 13
    3.3.2 EMSA ......................................................................................................................... 14
    3.3.3 MST ............................................................................................................................. 14
4. Results ....................................................................................................................................... 15
  4.1 Small Scale Expression Test ................................................................................................. 15
  4.2 Protein Purification ............................................................................................................... 16
    4.2.1 C9 Construct ................................................................................................................ 16
    4.2.2 C11 Construct .............................................................................................................. 18
    4.2.3 C13 Construct .............................................................................................................. 20
  4.3 Electrophoretic Mobility Shift Assay .................................................................................... 22
  4.4 Microscale Thermophoresis .................................................................................................. 26
5. Discussion .................................................................................................................................. 28
6. Conclusion ................................................................................................................................. 32
7. Acknowledgements .................................................................................................................... 33
8. References .................................................................................................................................. 34
Abstract

The zinc finger protein, zBED6, is a transcriptional regulator of IGF2 along with hundreds of other genes relating to development and growth. Studies on the growth of commercially bred pigs discovered a single nucleotide substitution in the third intron of IGF2 which disrupts the binding of zBED6 and is responsible for the three-fold upregulation of IGF2 in skeletal muscle. The mutation is linked to decreased subcutaneous fat deposition, larger organ size, and increased skeletal muscle mass. Three different constructs of the zBED6 protein made by Björklund 2018 were expressed and purified to characterize their binding affinity, where one contained both zinc finger domains and two of the constructs contained only one zinc finger domain each. Electrophoretic mobility shift assay protocol was optimized to determine the apparent $K_a$ ($= 210 \pm 31 \text{nM}$) for the full-length construct C13 and to determine which zinc finger domain was sensitive to the mutation in the IGF2 gene. The first zinc finger domain seems to be more specific in its binding target. Preliminary microscale thermophoresis results were highly variable, needing further optimization of the protocol in order to obtain a full binding curve. The next steps involve site directed mutagenesis of residues binding DNA to determine which interactions are the most significant and possibly crystallization studies as well.
List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>zBED6</td>
<td>Zinc finger BED-type containing protein 6</td>
</tr>
<tr>
<td>IGF2</td>
<td>Insulin-like growth factor 2</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribose nucleic acid</td>
</tr>
<tr>
<td>ChIP-Seq</td>
<td>Chromatin immunoprecipitation – DNA sequencing</td>
</tr>
<tr>
<td>EMSA</td>
<td>Electrophoretic mobility shift assay</td>
</tr>
<tr>
<td>MST</td>
<td>Microscale thermophoresis</td>
</tr>
<tr>
<td>ITC</td>
<td>Isothermal titration calorimetry</td>
</tr>
<tr>
<td>QTL</td>
<td>Quantitative trait loci</td>
</tr>
<tr>
<td>IMAC</td>
<td>Immobilized metal ion affinity chromatography</td>
</tr>
<tr>
<td>SEC</td>
<td>Size exclusion chromatography</td>
</tr>
<tr>
<td>LB</td>
<td>Lysogeny broth</td>
</tr>
</tbody>
</table>
1. Introduction

1.1 The zBED6 transcription factor
The domestication and subsequent selective breeding of wild pigs has resulted in an increase in skeletal muscle mass and reduced deposition of subcutaneous fat, without negatively affecting birth weight and growth. These complex traits are commonly thought to arise from interactions between two or more genes and their environment (Markljung et al. 2009). The underlying genes responsible for these quantitative traits can be linked to stretches of DNA known as quantitative trait loci (QTL). A paternally expressed QTL, mapping to a region in the insulin-like growth factor 2 gene region (IGF2), affects muscle development, organ growth, and fat deposition (Van Laere et al. 2003; Younis et al. 2018). The causative mutation of this QTL is a single nucleotide substitution (G > A) in the third intron of IGF2, occurring in a conserved CpG island in placental mammals (Van Laere et al. 2003). The mutation disrupts the interaction of zBED6, a nuclear protein acting as a repressor of IGF2 transcription, resulting in the upregulation of IGF2 mRNA in skeletal muscle by three-fold (Van Laere et al. 2003; Markljung et al. 2009).

The repressor interacting with the IGF2 gene region was identified by an affinity capture method using biotinylated oligonucleotides followed by mass spectrometry (Markljung et al. 2009; Andersson et al. 2010). Nuclear extracts from C2C12 mouse myoblasts were labelled and used to find proteins which bound wild-type but not mutated oligonucleotides. The protein demonstrating the highest enrichment by wild-type oligos corresponded to an annotated transcript of an alternative splice form of the Zc3h11a gene. Zc3h11a belongs to a large family of zinc finger proteins. However, it was found that the captured peptide, named zBED6, was encoded by an intronless gene in intron 1 of the Zc3h11a gene and bears no sequence similarity to Zc3h11a (Markljung et al. 2009; Andersson et al. 2010).

zBED6 is a protein encoded by an exapted transposon, containing two N-terminal BED-type zinc finger domains and a C-terminal hATC dimerization domain, as seen in Figure 1 (Markljung et al. 2009); therefore, zBED6 is related to the hAT family of DNA transposons. The BED zinc finger domain was initially discovered through bioinformatic studies of chromatin boundary element binding proteins DREF and BEAF from Drosophila melanogaster. PSI-BLAST results revealed several proteins sharing a signature which was predicted to form some variant of a BED-type zinc finger (Aravind 2000; Hayward et al. 2013). zBED6 is highly conserved among placental mammals, with the DNA binding BED
domains showing near 100% sequence identity between 26 species; Additionally, the BED domains in zBED6 are more closely related to each other than to the domains of other zBED proteins (Markljung et al. 2009).

Figure 1: Schematic of zBED6, showing the BED zinc finger domains and the hATC dimerization domain (modified from Markljung et al. 2009).

Chromatin Immunoprecipitation – DNA Sequencing (ChIP-Seq) had identified a consensus sequence for zBED6 binding: 5′-GCTCGC-3′ which is conserved among eight mammalian species (Van Laere et al. 2003). Furthermore, ChIP-Seq experiments using anti-zBED6 antibody had been done to identify other potential binding targets of zBED6. Data analysis of 24 million reads aligned to the mouse genome revealed 2,499 peaks, representing sites interacting with zBED6. Approximately 50% of the sites were found downstream of transcription start sites, prompting the proposal of a transcriptional silencing mechanism, possibly by chromatin remodelling. Members of the zBED family, such as the Drosophila DREF protein, have been shown to form complexes with chromatin remodelling complex NURF. Its human ortholog, zBED1, interacts with MI2, a chromatin remodelling factor and PC2, a polycomb protein known to modify and induce structural change in chromatin (Markljung et al. 2009; Grossniklaus & Paro 2014).

While zBED6 was initially discovered as a repressor for IGF2, due to its broad tissue distribution and ChIP-Seq data, it is expected to be a regulator for hundreds if not thousands of other genes as well. Up to 1,200 annotated genes were associated with the putative zBED6 binding sites found and were located within 5kb of the IGF2 gene. These genes were found to be connected to the development and regulation of basic biological processes; transcription; cell differentiation and signalling; and muscle development (Markljung et al. 2009). Understanding how a mutation in IGF2 affects zBED6 binding will help our understanding of how hundreds of growth-related genes are regulated.
1.2 Zinc Finger Structure and Function

The zinc finger domain is normally 28-30 amino acids which forms a β-hairpin, an antiparallel β sheet, preceding an α-helical structure. A zinc ion is coordinated by two conserved cysteine residues located on the β sheet and two conserved histidine residues found at the C-terminus of the α-helix, hence the name Cys$_2$ – His$_2$ zinc finger (Razin et al. 2011; Fedotova et al. 2017). The interaction between the zinc, cysteine, and histidine residues stabilizes the protein fold, illustrated in Figure 2. Zinc fingers are involved in a broad range of functions, such as: transcription, translation, cell signalling, and apoptosis; therefore, they can interact with several types of molecules, including nucleic acids and proteins (Krishna et al. 2003).

Zinc fingers normally recognize the major groove of DNA, where a distinct sequence of electron donors and acceptors allow for specific recognition and binding. Recognition of specific DNA sequences is accomplished by the side chains on the α-helix structure (Fedotova et al. 2017). One of the most well characterized Cys$_2$ – His$_2$ zinc fingers is Zif268, a transcriptional repressor for genes relating to cell differentiation and mitogenesis. The crystal structure of Zif268 bound to DNA revealed that tandem zinc fingers bind the same nucleotides in consecutive major grooves using amino acids at the same α-helical locations. The first, second, and third nucleotides on the 5’-end are recognized by positions +6, +3, and

Figure 2: PyMoL image of PDB entry: 2DJR. The zinc ion coordinating the structure is seen in grey. The cysteine residues are highlighted in blue and the histidine residues in red.
-1, while position +2 interacts with a nucleotide on the complementary strand (Fedotova et al. 2017). Therefore, it should be possible to identify significant amino acids on other zinc finger proteins by using structural or sequence alignments.

There are features which distinguish BED zinc finger domains from canonical Cys2–His2 zinc fingers but these have only been briefly outlined based on predictive studies. Firstly, an N-terminal motif, with highly conserved aromatic amino acids is characteristic of BED zinc fingers. Secondly, a highly variable helical region is predicted between the N-terminal motif and the cysteine dyad. Finally, these proteins share a region enriched in basic residues N-terminal to their BED domains, which are thought to form contacts with the DNA minor groove (Aravind 2000).

1.3 Protein Purification

Functional and structural studies of protein normally require high levels, >95% pure after the final polishing steps. The most optimal purification schemes utilize techniques based on different separation principles. Additionally, there should be minimal sample handling between purification steps. It is preferable that the elution conditions for one step are suitable start conditions for the next.

1.3.1 Immobilized metal ion affinity chromatography

Immobilized metal ion affinity chromatography (IMAC) is a common first step when purifying proteins. Metal ions such as: zinc, nickel, copper, and cobalt ions immobilized on a column are frequently used to bind electron donor groups found on the surface amino acids of proteins. Compared to the target protein with a histidine tag, other molecules and impurities exhibit weaker binding to the metal ions are easily washed away. In this way, histidine-tagged fusion proteins coupled with IMAC has become an indispensable first step in the purification of proteins (Cheung et al. 2012).

1.3.2 Size exclusion chromatography

Size exclusion chromatography (SEC) is favourable polishing technique which uses size as its separation principle. SEC columns are packed with porous beads of dextran polymers. The polymer beads contain depressions and pores of different shape and size (Arakawa et al. 2010). Small molecules therefore, are retained in the column for longer and elute later than large molecules, as they have a larger accessible volume they can potentially occupy.
1.4 Methods to characterize DNA-protein interactions

1.4.1 Electrophoretic Mobility Shift Assay
Quantitative measurements of binding parameters are paramount in the characterization of specific interactions between DNA and proteins. There are several methods which allow for the determination of relative binding affinity and stoichiometry. One common method is the electrophoretic mobility shift assay (EMSA). The development of new imaging techniques has allowed for the use of a wider range of methods which rivals the sensitivity of $^{32}$P-labelled probes. Some of these methods include: fluorescence detection and chemiluminescence. Fluorescence detection uses commercially available dyes and is a good option when very high sensitivity is not required. Alternatively, a method involving biotin end-labelled DNA and a detection kit consisting of Streptavidin-Horseradish peroxidase and a chemiluminescent substrate has also been growing in popularity (Hellman and Fried, 2007).

EMSA works on the principle that: the electrophoretic mobility of a DNA-protein complex will be slower than that of free DNA, producing a visible shift at appropriate concentrations when run on polyacrylamide or agarose gels (Hellman and Fried, 2007). The shift in bands can then be quantified to estimate the equilibrium dissociation constant ($K_d$) (Heffler et al. 2012). The relationship between binding affinity and $K_d$ is inverse, where a low $K_d$ value indicates high binding affinity, whereas a high value suggests weaker binding to ligands (Heffler et al. 2012). For standard EMSA, a series of reactions is set up where the DNA concentration is below the $K_d$ and protein concentration ranges from below to above the $K_d$.

1.4.2 Microscale Thermophoresis
Microscale thermophoresis (MST) is an alternative approach for analysing a wide array of molecular interactions, from small molecule-protein interactions to quantifying the binding affinities of protein-protein and protein-nucleic acid interactions (Jerabek-Willemsen et al. 2011). Thermophoresis is defined as the directed flux of molecules induced by a temperature gradient (Jerabek-Willemsen et al. 2011). Samples are loaded into capillaries and subjected to heating by an infrared (IR) laser. The laser allows for high precision and reproducibility, which is important when a typical serial dilution analysis involves 10-16 samples (Jerabek-Willemsen et al. 2011; Seidel et al. 2013). The change in fluorescence of the labelled molecule over time reflects the distribution of molecules and is dependent on two effects. The
fluorescent intensity of a fluorophore will be affected intrinsically by changes to temperature. Additionally, the movement of molecules along a temperature gradient, coined thermophoresis, will affect fluorescence readings due to a change in molecule concentration. Furthermore, the extent of change in fluorescence will be altered by binding events, either by conformational change or by a ligand. A binding curve can then be derived from the thermophoresis curves, which allows for the calculation of binding affinity (Seidel et al. 2013).

2. Aims
The aim of the project is to further characterize the DNA-binding of zBED6 and answer the following questions:

- What is the affinity of ZBED6 to the wild type and mutated DNA binding site in the *IGF2* intron?
  - This will be tested using quantitative EMSA using fluorescently labeled DNA oligos.
  - Techniques involving MST will be explored as an alternative to derive equilibrium dissociation constant (*K*<sub>d</sub>)
- Which Zn finger is responsible for binding to which part of the DNA?
  - EMSA with single zinc-finger constructs will be tested for binding

In addition, we hoped to investigate which amino acids were involved in the interaction with the target sequence. This would have been done using by site directed mutagenesis coupled with EMSA.
3. Methods

3.1 Transformation of plasmid

Plasmids were constructed by Dennis Björklund, a previous master’s student in Maria Selmer’s group here at Uppsala University.

Table 1: zBED6 constructs and their respective truncations. C8 and C13 have both zinc finger domains. C9 only has the first zinc finger domain and C11 only has the second domain (adapted from Björklund 2018).

<table>
<thead>
<tr>
<th>Construct</th>
<th>Amino Acids</th>
<th>Construct Modifications</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>108-325</td>
<td>2 zinc fingers, N- and C-terminal truncations</td>
</tr>
<tr>
<td>9</td>
<td>92-209</td>
<td>Zinc finger 1 only</td>
</tr>
<tr>
<td>11</td>
<td>212-384</td>
<td>Zinc finger 2 only</td>
</tr>
<tr>
<td>13</td>
<td>92-384</td>
<td>2 zinc fingers with a small N-terminal truncation</td>
</tr>
</tbody>
</table>

3.1.1 Transformation to BL21(AI) and BL21(DE3) pLysS expression cells

First, 50μL of competent cells were thawed on ice for 30 minutes. 1μL of C13 and C8 plasmid was transferred to the tubes containing competent cells and left on ice for 30 minutes. The cells were then heat shocked at 42°C for 60 seconds and placed back on ice for 2 minutes. 150μL of SOC medium was added to the tube and incubated at 37°C on a shaker at 100rpm for 1 hour. 200μL was then spread on LA plates with 50μg/mL kanamycin and incubated overnight at 37°C. For transformation to BL21(DE3) pLysS cells, the plates also required 25μg/mL chloramphenicol.

3.1.2 Transformation to TOP10 competent cells and mini-prep

A 50μL vial of One Shot TOP10 chemically competent cells was thawed on ice for 30 minutes. 1μL of C13 plasmid was pipetted directly into the tube of cells and mixed by gentle tapping. The tubes were put back on ice for 30 minutes. Subsequently, the sample was heat shocked for 30 seconds at 42°C and placed on ice for 2 minutes. 200μL of SOC medium was added to the sample and incubated at 37°C at 100rpm for 1 hour. Then, 200μL was spread on LA plates with 50μg/mL Kanamycin and incubated overnight at 37°C.
A colony from the plate was grown in 10mL LB media with 50μg/mL kanamycin at 37°C 110 rpm overnight. The cells were harvested at 10,000 x g for 1 minute. Then, the plasmid was purified following the E.Z.N.A® Plasmid Mini Kit protocol. The plasmid was eluted with 100μL elution buffer supplied in the kit. The plasmid was sent for sequencing with primers pLIC_forw and pLIC_rev.

3.2 Protein Purification

3.2.1 Small Scale Expression Test
100mL of LB with 50μg/mL kanamycin was inoculated with a colony of C13 and C8 transformants from BL21(AI) and PLysS plates. The flasks were incubated at 37°C until OD_{600} reached 0.5 for the four cultures, at which point they were moved to an incubator and cooled at 18°C and induced at 0.7 OD_{600} with 0.2% L-arabinose and 0.5mM IPTG. 5mL of sample was taken before induction (BI) and stored at -20°C until SDS analysis.

BI samples were harvested at 4000 rpm and resuspended in lysis buffer (50mM Tris-HCl pH 7.5, 300mM NaCl, 5% glycerol, 5mM BME, 0.1% Triton-X100, 10mL Protease Inhibitor Tablet (Pierce)) and the cells were broken using a Sonicator VCX 130 (Sonics & Materials, USA) completing 3 cycles (10 seconds ON/ 30 seconds OFF). The sample was then centrifuged at 13000 rpm for 10 minutes to obtain lysate and cell pellet samples. Culture collected after induction was harvested in the same way. SDS was run according to standard protocol, described in its own section.

3.2.2 Large Scale Expression
A colony of BL21(AI) transformants of the C9, C11, and C13 constructs were grown overnight in 10mL LB medium with 50μg/mL kanamycin at 37°C on a shaker at 110rpm. 800mL of LB medium and 50μg/mL kanamycin was added to 2.8L shaker flasks. Each flask was inoculated with 5mL of overnight culture, two flasks per construct. The flasks were then left shaking at 37°C at 110rpm until OD_{600} reached approximately 0.5, where they were transferred to a Multitron Pro (Infors HT, Switzerland) and left to cool at 18°C and 100rpm. Once an OD_{600} of approximately 0.7 was obtained, induction was carried out with 0.2% L-arabinose and 0.5 mM IPTG and left overnight at 18°C and 100rpm. In later expression experiments, induction was only done with 0.2% L-arabinose.

Cell harvesting was done at 6000 rpm for 20 minutes at 20°C in the Avanti J-26S XP using the 9.1000 rotor (Beckman Coulter, USA). The resulting pellet was resuspended in 20mL
wash buffer (50mM Tris-HCl pH 7.5, 150mM NaCl) and centrifuged at 4000 rpm for 30 minutes at 8°C in the Allegra X-30R (Beckman Coulter, USA). The supernatant was poured off and the pellets were stored at -20°C until purification.

3.2.3 Protein Purification
For each purification, a lysis buffer cocktail consisting of: 50mL lysis buffer (50mM NaH$_2$PO$_4$ pH 7.5, 300mM NaCl, 10% glycerol), protease inhibitor tablet (Pierce™ 50mL tablets), 500μL of 10% Triton-X100, 1μg/mL DNase, and 5mM BME was made. Cell pellets (4-6 g) were resuspended in 20mL lysis buffer cocktail at 8°C using a magnetic stirrer.

Cell lysis 1: The cells were lysed using a cell disruptor at 0.98 kPa in the Constant Cell Disruption System (Constant System Ltd, UK).

Cell lysis 2: A Sonicator VCX 130 (Sonics & Materials, USA) was used on a cycle of 10 seconds ON/30 seconds OFF for as many cycles could be completed in 3 minutes.

Afterward, the cells were centrifuged at 16,000 rpm for 45 minutes at 4°C in the Sorvall™ RC6 (Thermo-Fisher Scientific, USA) using the SS-34 rotor. In this time, a 1mL bed volume of Ni-Sepharose in a gravity column was washed 3 times with 25mL of water and equilibrated with 10mL lysis buffer (50mM NaH$_2$PO$_4$ pH 7.5, 300mM NaCl, 10% glycerol). After centrifugation, the supernatant was collected and was passed through a 0.45μm filter into the gravity column. The column was incubated for one hour at 8°C on a tilting table; after, the column was run and the flow through was collected in a 50mL falcon tube. The column was washed using 15mL of wash buffer (50mM Tris-HCl pH 7.5, 1M NaCl, 5% glycerol, 5mM BME), then with 75mL wash buffer with 20mM Imidazole. The A$_{280}$ of the wash was measured in a nanodrop 2000 after every 15mL. The protein was eluted with 10mL elution buffer (300mM Imidazole, 50mM Tris-HCl pH 7.5, 300mM NaCl, 5% glycerol, 5mM BME) as 1mL fractions. The A$_{280}$ of the fractions were checked in a nanodrop 2000, discarding any fractions with values below 0.2. Optimal fractions were pooled and concentrated down to 4mL and run on the Superdex 200 16/60 (GE Healthcare, USA). The column was first equilibrated with filtered gel filtration buffer (Tris-HCl pH 7.5, 300mM NaCl, 5% glycerol, 5mM BME).

SDS-PAGE was run with peak fractions identified in the chromatogram as well as the: lysate, flow through, cell pellet, wash, and elution fractions.
3.2.4 SDS-PAGE Protocol
2μL of lysate, flow through, and eluate samples were mixed with 5μL of 5X Laemmli buffer and diluted with 8μL of distilled water. Cell pellet samples were mixed with 5μL 5X Laemmli buffer and diluted with 10μL water. For the remaining fractions, 10μL of sample was taken and mixed with 5μL 5X Laemmli buffer. Samples were boiled at 95°C for 10 minutes and spun down briefly. Samples were loaded into a Mini-PROTEAN® TGX Stain-Free pre-cast gel (Bio-Rad, USA) and run at 200V in 1X running buffer (25mM Tris, 192mM glycine, 0.1%SDS). The Precision Plus Protein™ Dual Color Standards (Bio-Rad, USA) was used as a ladder.

3.3 Electrophoretic Mobility Shift Assay

3.3.1 Annealing of Double-Stranded DNA
Wild-type (WT) and mutant (Mut) dsDNA targets of the zBED6 protein was made using the following forward and reverse oligos:

<table>
<thead>
<tr>
<th>DNA</th>
<th>Forward</th>
<th>Reverse</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>5'-AGA TCC TTC GCC TAG GCT CGC AGC GCG GGA GCG A-3'</td>
<td>5'-TCG CTC CCG CCG TGC GAG CCT AGG CGA AGG ATC T-3'</td>
</tr>
<tr>
<td>Mut</td>
<td>5'-AGA TCC TTC GCC TAG GCT CAC AGC GCG GGA GCG A-3'</td>
<td>5'-TCG CTC CCG CCG TGT GAG CCT AGG CGA AGG ATC T-3'</td>
</tr>
</tbody>
</table>

The following annealing protocol was run in the T100™ Thermal Cycler (Bio-Rad, USA):

<table>
<thead>
<tr>
<th>Cycles</th>
<th>Temperature (°C)</th>
<th>Minutes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>95</td>
<td>5</td>
</tr>
<tr>
<td>40</td>
<td>95 (-1°C/cycle)</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>55</td>
<td>15</td>
</tr>
<tr>
<td>20</td>
<td>55 (-1°C/cycle)</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>Hold</td>
</tr>
</tbody>
</table>

Table 2: dsDNA annealing protocol (adapted from Björklund, 2018)
3.3.2 EMSA
Preparation 1: A master mix containing 44μL 5X binding buffer (0.32% NP-40, 37.5% glycerol, 150mM KCl, 10mM MgCl₂, 6.5mM DTT, 10mM Spermidine, 75mM Heps-KOH pH 7.65), 33μL 1μM dsDNA, and 91.μL distilled water was made. The mix was evenly distributed between 10 PCR tubes, with one tube being a negative control (no protein or competitor DNA). Protein concentrations ranged from 0-200nM, increasing at 25nM increments. 0.0175μg/μL of Poly dI-dC unlabelled competitor DNA was added and distilled water was added to adjust the final volume to 20μL.

Preparation 2: A dilution series was set up to test protein concentrations ranging from 0-800nM with 100nM dsDNA. 200nM WT and Mut dsDNA was made in 1X binding buffer with 0.0175 μg/μL competitor. 10μL of DNA and protein were then mixed together in a PCR tube.

2.5% agarose gel was prepared by adding 50mL 0.5X Tris Borate (TB) buffer to 1.25g agarose and dissolved in a microwave. The gel was cast in a 7 x 8.5cm chamber and two 10 (or 12 for Preparation 2) well combs were inserted. The agarose was allowed to cooled for at least 20 minutes at room temperature (22°C). The samples were run at 130V for 15 minutes at room temperature.

The samples were spun down briefly and incubated at room temperature for 20 minutes before loading onto the gel.

The gel was imaged directly using a ChemiDoc MP (Bio-Rad, USA) under the nucleic acid “fluorescein” protocol. ImageLab was then used to quantify the shift of fluorescent bands using the well containing just DNA as the reference band. Non-linear regression was done using Excel solver (Brown 2001; Heffler et al. 2012)

3.3.3 MST
Using a 396 well plate, 20μL of 50μM protein was added into the first well. 10μL of assay buffer (Tris-HCl pH 7.5, 300mM NaCl, 5% glycerol, 5mM BME) was added into wells 2-12. A 1:1 serial dilution was done by transferring 10μL from well 1 to well 2 and mixing. This was repeated, and 10μL was discarded from the last well after mixing. 10μL of fluorescently labelled 60nM dsDNA was pipetted into each well and mixed. The plate was incubated at room temperature for 5 minutes before running in the Monolith NT. Automated (Nanotemper). MST experiments were run using 40% MST power and 60% excitation power.
4. Results

4.1 Small Scale Expression Test

A small-scale expression test was done in two different strains, BL21(AI) and pLysS. The test expression was done to determine if there was any difference in expression levels between the two cell strains before and after induction. Pellet and lysate fractions were taken to determine whether the constructs C13 and C8 were soluble. SDS-PAGE of lysate and pellet fractions before and after induction, as seen in Figure 3, showed bands in the expected location for each of the two constructs. Furthermore, the gel indicated that the C8 construct significantly less soluble, with most of the protein being found in the pellet fraction after induction.

Figure 3: SDS-PAGE of lysate and pellet fractions before and after induction of C8 from BL21(AI) cells and C13 from BL21(AI) and PLysS cells. The expected sizes of the proteins were 47.4 and 39.6 kDa for C13 and C8 respectively.

SDS-PAGE of protein purification fractions, seen in Figure 4, shows distinct bands in the cell pellet fraction for the C8 construct. Expression of C13 between BL21(AI) cells and pLysS cells did not seem to differ significantly. Several bands in the appropriate size range can be found in the elute fraction, but it is likely that very little to none of the protein of interest was eluted from the IMAC column. In contrast, bands around the expected size can be seen in the C13 elute column in Figure 4. However, the protein is likely less than 10% of what was eluted, indicating an additional purification step is needed.
4.2 Protein Purification

4.2.1 C9 Construct

The C9 construct was purified using a Ni-Sepharose gravity column and SEC was done using the Superdex 75 (Bio-Rad, USA) on the NGC Chromatography System (Bio-Rad, USA). SDS-PAGE of the C9 construct, Figure 6, highlighting the purification fractions shows loss of a significant amount of protein in the imidazole wash steps. The second wash, with buffer containing 40mM imidazole only seems to remove a few of the impurities while eluting more of the target protein than intended. The A12 fraction, Figure 5, obtained after SEC yielded relatively pure protein at a concentration of 0.721mg/mL before concentrating. The final concentration of C9 was 3.707mg/mL from 800mL culture. The A22 fraction was a distinct secondary peak seen after size exclusion.
Figure 6: SDS-PAGE of the C9 construct after IMAC and SEC purification steps. The predicted size of C9 is 28.5 kDa. Wells from the right show: lysate, pellet, flow through, Imidazole wash 1, Imidazole wash 2, elute, A12 and A22 (size exclusion chromatography fractions from Superdex 75 16/60) fractions. A12 corresponded to the peak fraction after SEC where the C9 construct was expected to be found (28.5 kDa).
4.2.2 C11 Construct

The protocol was modified by removing the wash step with 40mM Imidazole, resulting in less protein being eluted early while maximizing the amount of impurities washed out, seen in Figure 8. While there is a prominent band slightly above where the protein was expected on the SDS-PAGE gel, the SEC peak fractions corresponded well to where C11 was predicted to elute, Figure 7; therefore, the higher band location on SDS gel likely corresponds to the target protein. The protein was not purer after SEC because the peak fractions were pooled and concentrated without an SDS-PAGE, fractions with higher impurities ended up in the final sample. Overall C11 yield was low compared to the other constructs, 0.286mg/mL from 800mL culture, was obtained after concentrating peak fractions from SEC.
Figure 7: Chromatogram of C11 construct during SEC using the Superdex 75. The fractions where the protein was the peak ranging from A10-A12. These fractions were taken and run on SDS-PAGE. The first peak corresponds to protein aggregates while the last peak is likely non-target low molecular weight proteins.
4.2.3 C13 Construct

Due to the wide middle peak seen in Figure 9, several fractions after SEC were run on SDS-PAGE to see which should be pooled and concentrated for further use, Figure 10. Fractions B10 and B11 were saved due to their relatively high purity. Final concentrations of these fractions were 10.154mg/mL (207μM) and 4.324mg/mL (91.2μM) from 800mL culture respectively in approximately 40μL.
Figure 9: Chromatogram of C13 construct during SEC. The fractions where the protein was expected was a wide peak, ranging from B12-B6. These fractions were taken and run on SDS-PAGE. The first peak corresponds to protein aggregates while the last peak is likely non-target low molecular weight proteins.
Figure 10: SDS-PAGE of C13 construct after IMAC and SEC purification steps. The predicted size of C13 is 47.4 kDa. Wells from the left show: Pellet, flow through, wash, imidazole wash 1, imidazole wash 2, elute, B6-B12 (from size exclusion chromatography fractions from the Superdex 200) fractions.

4.3 Electrophoretic Mobility Shift Assay

To characterize the binding affinity of the various protein constructs to target DNA, EMSAs were run with a constant DNA concentration of 100nM with protein concentrations ranging from 0-800nM for WT and 0-1200nM for Mut conditions. The gel shift could be used to estimate the $K_d$ of the construct since there is a case where half the DNA has shifted. The binding of C13 to the WT dsDNA sequence was much stronger than its binding to the mutant sequence, as seen in Figure 11.
Figure 11: EMSA with C13 construct, WT DNA, and Mut DNA. Samples were run in 2.5% agarose gel at 130V for 15 minutes. Protein concentrations range from 0-800nM, labelled above the wells. The DNA concentration was constant, 100nM, in both experiments. 0.0175μg/μL Poly dI-dC unlabelled DNA was added to the DNA mix.

Preliminary tests of the C9 and C11 constructs with a wider range of protein concentrations, seen in Figure 12, indicate saturated binding at very high protein concentration with WT dsDNA. For both constructs, approximately 50% of the DNA seems to be bound at 200nM protein, so the same titration was used in the scaled up EMSA (Figures 13 and 14). Additionally, the gel shift in Figure 12, indicates C11 is less sensitive to the mutation due to the more prominent shifting and smearing of DNA across all protein concentrations.
Figure 12: EMSA with C9 and C11 constructs. The set of experiments on the left are with 100nM WT dsDNA. Samples were run in 2.5% agarose gel at 130V for 15 minutes. The experiments on the right are cases with 100nM Mut dsDNA. Protein concentrations in nM are labelled above the wells. 0.0175μg/μL Poly dI-dC unlabelled competitor DNA was added to the DNA mix.

The scaled-up experiments of C9 and C11, illustrated by Figures 13 and 14 respectively, only show very weak binding to the WT dsDNA. The shift for C9 from 100-800nM while the shift for C11 is only visible from 400-800nM. No binding can be seen with the mutant sequence, even at 1200nM protein concentrations.
Figure 13: EMSA of C9 construct with 100nM WT and Mut dsDNA. Samples were run in 2.5% agarose gel at 130V for 15 minutes. Protein concentrations in nM are labelled above the wells. 0.0175μg/μL Poly dI-dC unlabelled competitor DNA was added to the DNA mix.

Figure 14: EMSA of C11 construct with 100nM WT and Mut dsDNA. Samples were run in 2.5% agarose gel at 130V for 15 minutes. Protein concentrations in nM are labelled above the wells. 0.0175μg/μL Poly dI-dC unlabelled competitor DNA was added to the DNA mix.
The fraction of DNA bound was quantified using the well containing just DNA as the reference band for fluorescence. The fraction bound was plotted against C13 concentration (nM) as seen in Figure 15. The data were fit according to the Bolztmann equation (Brown 2001) and a $K_d$ was predicted at $210 \pm 31\text{nM}$. The $R^2$ value is 0.979.

![Figure 15: Scatterplot of fraction of DNA bound against C13 concentration (nM). Bands from Figure 9 were quantified in ImageLab and non-linear regression analysis was done in Excel using the solver plug-in (Brown 2001).](image)

4.4 Microscale Thermophoresis

MST was explored as an alternative method to quantify the binding affinity between the zBED6 constructs and target DNA and to corroborate results obtained from EMSA. The protein was incubated with 30nM WT and Mut DNA, which was kept constant. Experiments were run with the C13 construct (highest concentration $5\mu M$) resulted in a partial binding curve, as seen in Figure 16. An increasing MST signal can be seen with increasing protein concentration, $F_{\text{norm}}[\%]$ from 927 to 932 in the WT DNA case and from 926 to 930 in the Mut case.

Further experiments were done with C13 (highest concentration $25\mu M$) incubated with a fixed concentration of 30nM WT DNA only, to ensure there was saturated binding. An incomplete negative thermophoresis curve was obtained, seen in Figure 17. A decreasing MST signal was observed with increasing protein concentration, $F_{\text{norm}}[\%]$ from 932 to 917.
Figure 16: MST of C13 construct with WT DNA and Mut DNA. The highest protein concentration was 5μM and DNA concentration was constant at 30nM. Excitation power was set at 60% and MST power was set to 40%. Samples were incubated for 5 minutes at room temperature before measurements began.

Figure 17: MST of C13 construct with 30nM WT dsDNA. The highest protein concentration was 25μM. Excitation power was set to 60% and MST power was 40%. Samples were incubated for 5 minutes at room temperature before measurements began.
5. Discussion

The most amount of time in this project was spent during the expression and protein purification steps of the investigation. Initially, there was difficulty in obtaining good amounts of pure protein; therefore, small-scale expression tests had to be done using pLysS and BL21(AI) cells to compare outcomes. However, a successful protocol was devised for purification of the zBED6 constructs. Consequently, there was less time to conduct comprehensive binding assay tests, yet an apparent $K_d$ was obtained from EMSA experiments. MST was briefly explored but the experimental procedure will need further optimization.

The optimization of protein expression was key for performing protein-DNA binding assays. The small-scale expression and test purification, described by Figures 3 and 4, revealed that the C8 construct was less soluble than the C13, evident by the large bands seen in the pellet fractions. C8 was initially tried because the first expression and purification tests of the C13 construct was difficult. During the small-scale expression test, Figure 3, the cultures were induced slightly earlier than $OD_{600}$ 0.7, resulting in distinct bands where the protein was expected. The faintness of the bands in the image is a result of a lower volume of sample loaded onto the SDS-PAGE gel, it is likely that a more pronounced difference would be observable between BL21(AI) and PLysS if a greater volume of sample was used. Ultimately, it was decided that the protocol developed by Björklund 2018 would be used growth, where expression would be done using BL21(AI) cells and that induction had to be carefully timed with $OD_{600}$ 0.7. Induction of later cultures using 0.2% arabinose without 0.5mM IPTG did not seem to impact yields after protein purification. An autoinduction system, proposed by Sivashanmugam et al. 2009 could be an option to minimize handling of the cultures and samples. This method also seems to result in very good yield of protein at high cell densities.

Protein purification was successful following consecutive refinement attempts. Figures 9 and 10 reflect a more optimized purification protocol, resulting in highly pure C13 fractions following IMAC and SEC. Firstly, the 40mM imidazole wash illustrated in Figure 6 under the heading Wash 2 was removed for subsequent purifications as it caused the loss of substantial amounts of protein with little improvement to purity. Washing only included Wash buffer + 20mM imidazole, but the wash volume was increased to 75mL from 30mL. The wash fractions were checked on the nanodrop, after 15mL, to make sure the target
protein was not being lost. Prior to the modification to the washing step, ion exchange chromatography was considered as an extra purification step but would trade yield for higher purity. Furthermore, running the peak fractions identified from the size exclusion chromatogram on an SDS gel before pooling and concentrating the sample helps improve final purity of samples. The final C11 samples, Figure 8, suffer from low purity because all the fractions were pooled immediately after SEC.

The annealing of dsDNA strands has been somewhat variable throughout this research, as illustrated by the inconsistencies across EMSA gel images. It was posited that perhaps the agarose gel itself might affect the movement of DNA out of the well during electrophoresis, which causes blurry or smudged bands.

In order to obtain good results from EMSA, properly annealed DNA and protein pure enough to be confident regarding its concentration, is required. A clear separation of bound and unbound DNA should be visible following electrophoresis. Quantification of fluorescence from C13 EMSA, Figure 11, allowed for an estimation of binding affinity through nonlinear regression analysis. Following the nonlinear regression, it would be possible to refine the fit of the binding curve. Lowering the titration increments even further would be a suitable next step. However, it is recommended that the DNA concentration be as low as possible, far below the $K_d$; otherwise the protein concentration where half the DNA is bound would not be a suitable proxy of the $K_d$ (Heffler et al. 2012). 30nM DNA concentrations were tested by EMSA (data not shown) and no shift was detectable, even for the full C13 construct. Protocols using polyacrylamide gels have been successful in detecting sub-micromolar concentrations of bound fluorescently labelled DNA (Kim and Pabo 1998; Renda et al. 2007; Heffler et al. 2012). EMSA run on agarose gel however, can be run in a much shorter period. An ideal protocol might combine an agarose EMSA, to quickly identify protein concentration parameters, and native PAGE to solve apparent $K_d$.

The $K_d$ ($= 210 \pm 31\text{nM}$) of C13 derived from Figure 15 to other transcription factors and zinc finger proteins, is high in comparison. For example, the Gal4-p53 transcription factor has an apparent $K_d$ of $2.0 \pm 0.8\text{nM}$ (Heffler et al. 2012). The $K_d$ of zinc finger protein Zif268 was determined to be approximately $14 \pm 4\text{pM}$, indicating very high affinity toward its DNA target (Kim and Pabo 1998). One plausible explanation for this lower affinity is reflected in zBED6’s broad cellular distribution and its regulatory activity of hundreds of
other genes in addition to IG2. Conversely, refining of experimental parameters even further could likely obtain a more accurate estimation of binding affinity.

To determine which zinc finger was responsible for recognition of the mutation site, the C9 and C11 constructs were tested under similar conditions. Initially, EMSA with a wide range of protein concentrations revealed that both individual zinc fingers bound to the WT DNA sequence, highlighted in Figure 12. Furthermore, there was a prominent shift at 1600nM C11 incubated with mutant dsDNA. This prompted the idea that the C11 construct, having only zinc finger 2, was much less sensitive to the mutation at the binding site. However, the larger assays, represented by Figures 13 and 14 do not provide much more information. The shift for both constructs with the WT dsDNA was less pronounced in this test and there was no shift with the mutant dsDNA. It was expected that between 800-1600nM, C11 would still interact with the mutant sequence. Consequently, the preliminary EMSA might have been influenced by an error in setting up the appropriate binding conditions or variation between protein batches used in these experiments. Currently, it is difficult to state whether zinc finger 1 plays a larger role in recognizing the target DNA sequence. Renda et al. 2007 found similar results with an 11-zinc finger protein, where only a few domains were responsible for tight binding.

Preliminary experiments using MST resulted in highly variable outcomes between tests. A substantial shift in fluorescence should be at least 10 units; whereas the curve shown in Figure 16 is miniscule and does not result in a binding curve. Increasing the protein concentration to 25μM C13 also failed to attain a full binding curve; moreover, the partial binding curves are opposite shapes, suggesting an increase in thermophoretic mobility at low protein concentration (in this case 5μM) and a decrease in mobility at high concentrations (Fisher et al. 2017). It is probable that protein aggregation at this concentration has altered the fluorescence measurements and has been identified as a common problem with MST. However, a report by Rainard et al. 2018 suggests that the same DNA-protein complexes can produce both a positive thermophoretic shift, Figure 16, as well as a negative shift, Figure 17.

Increasing the protein concentration further is unlikely to make a difference in the binding curve. The $K_d$ values obtained from MST are significantly different from the value derived from EMSA. It is more likely that the $K_d$ for the WT obtained from the test using 5μM protein ($K_d = 1.21μM$) could be somewhat accurate. The fact that the experiment using Mut DNA did not obtain a $K_d$ value that was much higher than that of the WT DNA
experiment indicates that the protocol needs to be further optimized. Without a complete binding curve, one cannot derive a reasonable $K_d$. The required protein concentrations needed would end up being too high for replicate tests, which is why only the WT DNA was tested against higher concentrations of protein. Increasing the protein concentration to a maximum of $25\mu$M is likely too high to accurately derive $K_d$ from; however, this setup achieved a better MST response compared to the low protein setup, as the change in signal was greater than 5 units per the guidelines set by Nanotemper. Additionally, since the DNA is labelled instead of the protein, it becomes impossible to deduce whether the protein is adhering to the capillary, which can also affect the outcome. The addition of 0.05% Tween or any other detergent to the assay buffer would be a good counter-measure. Future experiments can look to optimize the MST protocol by modifying probes and reaction buffers. It is likely that an improved protein purification method coupled with an amended MST protocol can allow future experiments to obtain better results using less protein sample. An alternative method would be to try isothermal titration calorimetry (ITC) to derive binding affinity.
6. Conclusion

Good yield and purity of protein constructs C9 (only zinc finger 1) and C13 (both zinc fingers) was obtained following IMAC and SEC. EMSA results suggest the full protein has a binding affinity in the nanomolar range. MST was attempted to validate the quantification of EMSA data but the $K_d$ values obtained are likely inaccurate representations of the binding affinity of zBED6. A $K_d$ model is normally applied by the Nanotemper analysis software, giving a $K_d$ estimate; however, the model cannot be applied to the curves in Figures 16 and 17 as these would be highly inaccurate.

The protocol for expression and purification of protein has been modified and slightly optimized for purer proteins following IMAC and SEC, which can then be applied to the purification of C11 (only zinc finger 2) to obtain similar levels of purity to the other constructs. Ion exchange chromatography is still a potential alternative if yield is not an issue and higher purity is desired.

It is probable that the first zinc finger is more involved with specific binding, but further experiments are needed to corroborate this idea. Preliminary MST results are highly variable and can utilize a large volume of sample. Optimization of the procedure may allow lower sample volume use as well as improve the quality of the binding curves, but other methods such as ITC might be favourable.

Future work with zBED6 can involve the site directed mutagenesis of certain residues in the zinc finger domains to determine which residues are significant for specificity. ITC can be attempted to determine the affinity of constructs C9 and C11, because a method that is both highly sensitive and reproducible is needed. Further optimization of protein purification and expression protocols might allow for structural studies involving crystallization.
7. Acknowledgements

Throughout the writing process of this thesis I received an endless amount of support from those involved. First, I would like to thank Maria Selmer, whose expertise and leadership was invaluable in devising the investigation and its methodology. Furthermore, I would like to thank Sandesh Kanchugal for his guidance within the lab and for always being available to lend a hand.

I would like to extend my thanks to Annette Roos for showing me the methodology behind MST and the patience you showed during this learning period.

I would like to acknowledge Francisco Marcos-Torres for agreeing to be my defense opponent on such short notice.

Finally, an extended thank you to all those in the structural biology lab for subverting my expectations by making this experience engaging as well as memorable.
8. References


