IDEAL DNA BINDING SITE SPACING BY NF-KB P50 HOMODIMER

A Thesis
Presented to the
Faculty of
San Diego State University

In Partial Fulfillment
of the Requirements for the Degree
Master of Science in Biology
with a Concentration in
Molecular Biology

by
Christy Jane Milani
Summer 2015
SAN DIEGO STATE UNIVERSITY

The Undersigned Faculty Committee Approves the

Thesis of Christy Jane Milani:

Ideal DNA Binding Site Spacing by NF-κB P50 Homodimer

Tom Huxford, Chair
Department of Chemistry and Biochemistry

Greg Harris
Department of Biology

Thomas Cole
Department of Chemistry and Biochemistry

5/14/15
Approval Date
DEDICATION

To my parents, Parviz and Susan Milani, without whom none of my success would be possible.

To Dr. Tom Huxford, the best teacher I ever had, you inspire me every day.
I have no special talent. I am only passionately curious.
--Albert Einstein
ABSTRACT OF THE THESIS

Ideal DNA Binding Site Spacing by NF-κB P50 Homodimer
by
Christy Jane Milani
Master of Science in Biology with a Concentration in
Molecular Biology
San Diego State University, 2015

NF-κB is a family of eukaryotic transcription factor proteins that inducibly regulates the expression of a large number of genes that play important roles in innate immunity and inflammation. NF-κB consists of varying homo- and heterodimers, one form of which is a homodimer of p50 subunits (p50:p50). The NF-κB p50 homodimer is unique among NF-κB proteins in that it escapes regulation through association with classical IκB proteins in the cytoplasm and enters the nucleus where it binds to DNA. In the nucleus, p50:p50 is the preferred binding partner of the “nuclear IκB” protein IκBζ, itself an NF-κB responsive protein that is required for the further NF-κB-dependent expression of the pluripotent cytokine interleukin-6 (IL-6). This raises the possibility that ternary complexes of NF-κB p50 homodimer with IκBζ on target gene DNA could serve as a necessary signal for elevated expression of vital pro-inflammatory NF-κB target genes. Previously determined x-ray crystal structures of p50:p50 on κB DNA have revealed that this dimeric transcription factor binds to double-stranded DNA with variable spacing. We recently crystallized and determined the x-ray crystal structure of the DNA binding Rel homology region of murine p50 homodimer to κB DNA from the promoter of the IL-6 and NGAL genes and found, once again, that p50 binds the two sites with distinct spacing. In an attempt to identify the “preferred” binding mode of p50 on κB DNA, we prepared oligonucleotides with an “idealized” sequence κB DNA designed to allow p50:p50 to bind with 9-12 base pair spacing. The 3.0 Å x-ray co-crystal structure of the complex suggests that p50:p50 binds κB DNA with 11 base pair spacing when afforded freedom to select its ideal binding mode. We conclude that this is the preferred spacing of p50:p50 on κB DNA.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ABSTRACT</strong></td>
<td>vi</td>
</tr>
<tr>
<td><strong>LIST OF TABLES</strong></td>
<td>ix</td>
</tr>
<tr>
<td><strong>LIST OF FIGURES</strong></td>
<td>x</td>
</tr>
<tr>
<td><strong>CHAPTER</strong></td>
<td></td>
</tr>
<tr>
<td>1 INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Background</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Regulation of NF-κB by IκB Proteins</td>
<td>2</td>
</tr>
<tr>
<td>1.3 NF-κB and IκB Proteins</td>
<td>2</td>
</tr>
<tr>
<td>1.4 NF-κB Protein Structure and DNA Binding</td>
<td>5</td>
</tr>
<tr>
<td>2 METHODOLOGY</td>
<td>8</td>
</tr>
<tr>
<td>2.1 Purification of DNA Oligos</td>
<td>8</td>
</tr>
<tr>
<td>2.2 Annealing dsDNA</td>
<td>9</td>
</tr>
<tr>
<td>2.3 Protein Expression</td>
<td>9</td>
</tr>
<tr>
<td>2.4 Protein Purification</td>
<td>9</td>
</tr>
<tr>
<td>2.5 Complex Formation and Crystallization</td>
<td>10</td>
</tr>
<tr>
<td>2.6 X-Ray Diffraction Experiments</td>
<td>11</td>
</tr>
<tr>
<td>2.7 Structural Determination</td>
<td>11</td>
</tr>
<tr>
<td>2.8 Model Building and Refinement</td>
<td>12</td>
</tr>
<tr>
<td>3 RESULTS</td>
<td>13</td>
</tr>
<tr>
<td>3.1 Protein Expression and Purification</td>
<td>13</td>
</tr>
<tr>
<td>3.2 DNA Oligomer Design and Purification</td>
<td>15</td>
</tr>
<tr>
<td>3.3 Crystallization &amp; X-Ray Diffraction</td>
<td>15</td>
</tr>
<tr>
<td>3.4 Ideal P50 Homodimer Structure</td>
<td>20</td>
</tr>
<tr>
<td>4 DISCUSSION</td>
<td>24</td>
</tr>
</tbody>
</table>
ACKNOWLEDGEMENTS .....................................................................................................28
REFERENCES ........................................................................................................................29
LIST OF TABLES

PAGE

Table 3.1. Data Collection and Refinement Statistics .........................................................21
LIST OF FIGURES

Figure 1.1. Schematic diagram of NF-κB activation pathway. Inactive NF-κB is present in the cytoplasm. After activation by inducing stimuli, the inhibitor, IκB, becomes phosphorylated, ubiquitinated, and then degraded by the proteasome. The now active NF-κB can translocate to the nucleus and bind to DNA. ..............................................................................................................................3

Figure 1.2. Mammalian NF-κB family subunits. Each of the five protein subunits that make up the NF-κB family contain the Rel Homology Region (RHR). This region consists of the amino terminal domain (NTD) and the dimerization domain (DD); joined by a short flexible linker and a carboxy-terminal region that contains the Nuclear Localization Signal (NLS). Members of the class II subunits also contain a transcription activation domain (Trinh et al., 2008). ................4

Figure 1.3. (a) Harvard (H) and Yale (Y) DNA target sites. (b) Cartoon of p50 homodimer domain movements on DNA. The light shading on the N-terminal domain indicates the Harvard complex and the dark shading on the N-terminal domain represents the Yale complex. ............................................................................6

Figure 2.1. Schematic of DNA oligomers tested. This diagram shows the oligomers that were purified and screened in crystallization studies. Oligomers contain ends with either a 1-bp overhang (OH) or blunt ends (BE). Highlighted in red are the Guanines that the two p50 subunits bound to. ...................................................8

Figure 3.1. Purification of NF-κB p50. (A) SDS-PAGE gel of ion exchange purified p50 protein. (B) Chromatograph from size exclusion chromatography using Superdex 200 26/60. The blue curve represents protein, which absorbs at 280nm. The x-axis is the elution volume in mL and the y-axis is absorbance at 280 nm in mAU. ..........................................................................................................14

Figure 3.2. DNA Purification of κB DNA Oligos. The x-axis corresponds to elution volume in mL and the y-axis is in mAU. (A) Chromatograph from size exclusion chromatography of 16mer OHT (B) Chromatograph of 16mer OHA DNA Purification. ........................................................................................................16

Figure 3.3. Crystallization of p50:Ideal DNA complexes. (A) Crystal obtained for p50:16mer OH complexes grown in 4% v/v Tacsimate at pH 6.0 and 12% w/v Polyethylene Glycol. The scale bar equals 200 microns. .......................................................17

Figure 3.4. Diffraction Image. The crystal was tested for diffraction at the Advanced Photon Source station 24- ID- beam C at Argonne National Laboratory.
Shown here is the crystal data diffraction pattern with XDS indexed intensities used to determine the structure factors. ..........................................................18

Figure 3.5. Ribbon diagram of p50 (39-363) Rel Homology Region. (A) p50 subunits are represented in green. The multi-colored barrel in the middle represents the DNA. The C represents the dimerization domain and the N represents the DNA binding domain. (B) The view rotated 90°..........................................................22

Figure 3.6. Protein-DNA interactions of the p50 homodimer bound to various promoters. From left to right, the Yale structure of p50 homodimer bound to a symmetrized κB site and 12 base pair spacing between the p50 subunits, Harvard structure bound to MHC class I enhancer site and 11 base pair spacing, p50 bound to IL-6 promoter and 11 base pair spacing, p50 bound to the NGAL promoter and 12 base pair spacing, and the structure I solved with the p50 homodimer bound to “idealized” κB DNA and 11 base pair spacing. The amino acids histidine, and two arginine’s are labeled on where they contact the DNA promoters. ..................................................................................23

Figure 4.1. This is the electron density of the DNA with p50 and without p50. .....................25
Figure 4.2. Electron density zoomed in on p50: DNA contacts. Three adjacent guanines are bound to histidine, arginine, and arginine. The carbon atoms are represented by the color yellow, blue represents nitrogen, and pink represents oxygen..............................................................26
Figure 4.3. Crystal structure of p50 homodimer bound to “ideal” κB DNA turned 90° to the left. ............................................................................................................27
CHAPTER 1

INTRODUCTION

1.1 BACKGROUND

NF-κB is an inducible dimeric transcription factor that controls the expression of inflammatory and cell survival genes (Baeuerle, 1991). NF-κB was first discovered by David Baltimore’s lab in 1988 (Baeuerle, 1988a). By using mobility shift assays, they found one factor that bound to the κ light chain enhancer with the sequence 5'-GGGACTTTCC-3'. They named the new protein NF-κB because it is a nuclear factor that binds to the κ light chain enhancer region of activated B cells (Ghosh et al., 2012). Subsequent experiments showed that NF-κB is, in fact, found in the cytoplasm of all mammalian cells in an inhibited state (Baeuerle, 1988a). Treatment of cells with bacterial lipopolysaccharide (LPS), phorbol esters, or inflammatory cytokines caused NF-κB to become activated and transcription of target genes such as the immunoglobulin κ light chain gene to begin (Baeuerle, 1988b). NF-κB induces the expression of secreted cytokines, cell adhesion molecules, and cytoplasmic proteins that block apoptosis pathways to give the cell a better chance to survive (Ghosh and Hayden, 2008). Constitutive NF-κB activation is associated with inflammatory diseases such as Arthritis, Multiple Sclerosis, and Asthma (Courtois and Gilmore, 2006).

Dr. Patrick Baeuerle, another member of the Baltimore group, discovered that the cytoplasmic form of NF-κB was maintained inactive through its noncovalent association with an inhibitor protein. Dr. Baeuerle found that cytoplasmic NF-κB could be released from the inhibitor protein in vitro by treatment with a nonionic detergent called sodium deoxycholate, thus allowing NF-κB to bind DNA (Baeuerle, 1991). This allowed him to purify the inhibitor, which he named IκB (Baeuerle and Baltimore, 1989).
1.2 Regulation of NF-κB by IκB Proteins

NF-κB is present in the cytoplasm in association with inhibitor IκB proteins. After activation by an inducer like LPS, the IκB proteins become phosphorylated, ubiquitinated, and then degraded by the proteasome (Chen et al., 1995). The degradation of IκB allows NF-κB proteins to translocate to the nucleus by virtue of a type I nuclear localization signal (NLS) contained within the conserved N-terminal portion of the NF-κB subunits known as the Rel Homology Region (RHR); (Li and Verma, 2002). Once within the cell nucleus, NF-κB dimers bind to their cognate DNA binding sites and orchestrate the assembly of transcription factors and, ultimately, elevate the expression of target genes including cytokines, chemokines, and anti-apoptotic proteins (Chen et al., 1998).

IκBα plays an important role in the regulation of NF-κB (Figure 1.1). Previous studies have shown NF-κB to be continuously active in mice that were deficient in IκBα. Inducing stimuli such as a virus, bacteria, or inflammatory cytokine, causes the IκB kinase complex (IKK), made up of three proteins: IKKα/IKK1, IKKβ/IKK2, and IKKγ/NEMO, to become activated (Gilmore, 2006). The catalytic activity of the IKKβ subunit is responsible for phosphorylation of serine residues 32 and 36 on IκBα, thus targeting IκBα for ubiquitination via the β-TrCP1 SCF-type E3 ubiquitin-protein ligase complex, and degradation (Trinh et al., 2008). In the absence of IκB inhibitory proteins, cytoplasmic NF-κB translocates to the nucleus where it binds to target gene promoters and activates their transcription (Chen et al., 1995).

1.3 NF-κB and IκB Proteins

NF-κB is a protein and transcription factor that binds to DNA within the enhancer region of immunoglobulin κ light chain of B lymphocytes. NF-κB is part of a family of transcription factors that includes the p65, c-Rel, RelB, p50/p105, and p52/p100 subunits (Li and Verma, 2002). Functional NF-κB is composed of homo- and heterodimers of these proteins, all of which share a Rel homology domain (Figure 1.2). p105 is the precursor to p50 and p100 is the precursor of p52 (Li and Verma, 2002). Mature p50 and p52 subunits are generated from the incomplete proteolysis of p105 and p100 precursors, respectively, in a process that is regulated by IKK (Hirotani et al., 2005). Most of the NF-κB found in the cell
is in the form of a p65:p50 heterodimer, followed by the p50 homodimer, and the p65 homodimer (Li and Verma, 2002).

Figure 1.1. Schematic diagram of NF-κB activation pathway. Inactive NF-κB is present in the cytoplasm. After activation by inducing stimuli, the inhibitor, IκB, becomes phosphorylated, ubiquitinated, and then degraded by the proteasome. The now active NF-κB can translocate to the nucleus and bind to DNA.
The inhibitor of NF-κB, called IκB, is made up of many members including IκBα, IκBβ, IκBε, IκBζ, Bcl-3, p100, and p105 (Li and Verma, 2002). When NF-κB is in its inactive form in the cytoplasm, it is in complex with either the classical IκBα, IκBβ, and IκBε or with the NF-κB precursor proteins p105 or p100 (Hoffmann et al., 2006). All IκB proteins contain a common domain consisting of 6-7 ankyrin repeats. Ankyrin repeats are helical repeat motifs of supersecondary structure, each of which is roughly 33 amino acids in length (Kawakami et al., 1988). All the classical IκB proteins also contain a PEST peptide sequence, which stands for proline (P), glutamic acid (E), serine (S), and threonine (T) (Malek et al., 2001). This region of amino acids is common for proteins, such as IκB, that exhibit rapid turnover in the cell.
1.4 NF-κB Protein Structure and DNA Binding

In 1995, Dr. Gourisankar Ghosh and colleagues determined the x-ray crystal structure of the Rel Homology Region from mouse NF-κB p50 bound as a homodimer to an idealized palindromic κB target while in Paul Sigler’s lab at Yale. This structure shows the Rel Homology Region folds into 2 domains. The C-terminal domain or dimerization domain, containing a core β-sandwich structure, is responsible for dimer formation. The binding surface of the N-terminal domain has a variable length α-helical segment that forms strong contacts with the minor groove near the center of the κB element (Ghosh et al., 1995).

Around the same time, Christopher W. Müller and Stephen C. Harrison reported a p50 homodimer structure from their lab at Harvard. The two structures are very much alike, but the key difference between the two structures is the half site spacing, which results in different orientations of the two domains (Figure 1.3).

In the Harvard structure two arginine residues and a histidine from a conserved loop in the N-terminal domain bind three adjacent guanine bases on the 5' end of the major groove (Müller et al., 1995). There is also an A-A mismatch and a 5 and a half base pair half site spacing or 11 bp recognition site. The Yale researchers reported a 10 base pair spacing. However, this neglects the fact that if they had used a longer DNA, then histidine residues could have contacted two more bases at the ends of the DNA, suggesting that the DNA binding conformation observed in the Yale structure is really a 12 base pair recognition site with 6 base pair half site spacing.

Furthermore, the DNA two-fold symmetry axis in the Harvard complex bends 15° to either side of the central AAT; whereas the DNA in the Yale complex is straight (Müller et al., 1995). The difference in bending could possibly be one of the ways in which the complex accommodates the different half site spacing. So, which is the preferred orientation by which p50 binds DNA?
Figure 1.3. (a) Harvard (H) and Yale (Y) DNA target sites. (b) Cartoon of p50 homodimer domain movements on DNA. The light shading on the N-terminal domain indicates the Harvard complex and the dark shading on the N-terminal domain represents the Yale complex. Source: Müller, C.W., Rey, F.A., Sodeoka, M., Verdine, G.L., and Harrison, S.C. (1995b). Structure of the NF-kappa B p50 homodimer bound to DNA. Nature 373, 311–317.

IkBζ is a nuclear protein that shows sequence homology to cytoplasmic IkB proteins and binds with specificity to NF-κB p50 subunits. Interestingly, IkBζ is an inducible protein that is not expressed in resting cells. Stimuli that induce NF-κB activity, such as bacterial lipopolysaccharide (LPS) or interleukin-1 (IL-1), result in IkBζ expression and localization to
the nucleus. Macrophages derived from mice lacking the gene for IkBζ were found to be
incapable of expressing the pro-inflammatory cytokine interleukin-6 (IL-6) in response to
stimuli (Trinh et al., 2008). This suggests that IkBζ, itself an NF-κB-dependent gene,
functions to control the expression of other NF-κB-dependent genes such as IL-6. One
possibility is that by forming a complex with p50 homodimer at enhancer elements of
specific target genes, IkBζ functions as a transcriptional co-activator. A former student in the
lab named Norman Zhu recently solved the x-ray crystal structure where the p50 homodimer
Rel Homology Region bound to a 17-mer DNA target site from the IL-6 promoter which
contains the sequence 5′-GGGAAAATCCC-3′ (Trinh et al., 2008). This site has three
guanines for histidine, arginine, and arginine on both sides so 12 base pair binding (such as
observed in the Yale structure) was expected. However, the IL-6 complex has an 11 bp
recognition site like the Harvard complex.

Zhu also had a p50 homodimer structure with another IkBζ dependent promoter,
NGAL, which stands for neutrophil gelatinase-associated lipocalin, or sometimes called
Lipocalin 2. NGAL is involved in innate immunity by sequestrating iron that in turn limits
bacterial growth. The NGAL promoter caused p50 to adopt a 6 base pair spacing
conformation (Trinh et al., 2008). This is the same conformation seen in the structure from
Yale. This caused the N-terminus domains to move further apart, pushing histidine 64 of the
recognition loop off of the promoter. In the NGAL structure, the histidine 64 is hydrogen
bonding with the neighboring DNA stacked right on top of the bound DNA.

So, in summary, the NGAL and Yale structures have the same conformation (12 bp
recognition site). The IL-6 and Harvard structures have the same conformation (11 bp). It
remains unclear what drives the adoption of p50 to 12- or 11 bp site DNA binding or which
spacing is energetically preferred. In order to test whether NF-κB p50 homodimer prefers
binding with 11- or 12 bp spacing, we designed an “idealized” DNA target site that allows
several options for optimal NF-κB:DNA base contacts with different spacing. We have
crystallized the complex and determined its x-ray crystal structure. It is hoped that
understanding how the p50 homodimer recognizes and binds to different κB DNA target sites
might shed light on the mechanism by which IkBζ selectively regulates the expression levels
of particular NF-κB target genes.
CHAPTER 2

METHODOLOGY

2.1 PURIFICATION OF DNA OLIGOS

Individual oligos designed by Dr. Tom Huxford, were purchased from Integrated DNA Technology and received in lyophilized form (Figure 2.1). Oligos were resuspended in 800 μL of low salt buffer consisting of 0.01 M NaOH and 0.1 M NaCl. They were then centrifuged at 14,000 rpm for 5 minutes. The solution was then loaded onto a source 15Q anion exchange column (Pharmacia) equilibrated with 90% low salt buffer and 10% of high salt buffer consisting of 0.01 M NaOH and 1M NaCl. After loading, the column was then washed with 15 mL of 10% high salt buffer before being eluted using a salt gradient of 10-90% high salt buffer over 12 column volumes. 100 μL of 1M MES pH 5.8 was added to each fraction containing DNA and fractions were pooled. The pooled fractions were concentrated to a volume of 100 μL using an Amicon Ultrace-3 concentrator. Buffer exchange was performed, to lower the salt concentration, using 20 mM Tris pH 7.5 and solution was concentrated again.

14 mer OH 5’-GAG GGG AAT TCC CCT-3’
3’-TC CCC TTA AGG GGAC-5’
16 mer OH 5’-TCA GGG GAA TTC CCC TC-3’
3’-GT CCC CTT AAG GGG AGA-5’
16 mer BE 5’-CAG GGG AAT TCC CCTC-3’
3’-GTC CCC TTA AGG GGAG-5’

Figure 2.1. Schematic of DNA oligomers tested. This diagram shows the oligomers that were purified and screened in crystallization studies. Oligomers contain ends with either a 1-bp overhang (OH) or blunt ends (BE). Highlighted in red are the Guanines that the two p50 subunits bound to.
2.2 ANNEALING dsDNA

Equimolar amounts of complementary oligos were combined. The sample was then incubated in heat block at 100°C for 5 minutes. After this, samples were removed from the heat block and allowed to slowly cool to room temperature. Double stranded DNA was concentrated to 2-3 mM with the Amicon Ultracel-3 concentrator and stored at -20°C.

2.3 PROTEIN EXPRESSION

A pET11a expression vector containing cDNA insert of murine p50 (39-363) was constructed as described previously (Chen et al., 1998). The vector was then transformed into E. coli BL21(DE3) cells by heat shock method. The cells were then plated on LB/AMP agar plates. The next day a single colony was picked and mixed with a 4 mL LB/AMP(100μg/mL) starter culture at 37°C and 225 rpm for 3-4 hours to allow it to grow. 2 mL of starter culture was added to a two liter culture of LB media containing ampicillin (100μg/mL). The large culture was agitated at 225 rpm for 3-5 hours at 37°C. After this time period, the large culture was removed from the shaker and placed on a magnetic stir plate to cool to room temperature and stirred for one hour. Isopropylthiogalactoside (IPTG), with a concentration of 0.1 mM, was added to the culture to induce protein expression and left stirring at room temperature overnight for 16 hours. The culture was then centrifuged at 4200 rpm for 15 minutes, re-suspended in 1x PBS (Phosphate buffered saline) buffer, centrifuged again at 4200 rpm for 15 minutes, poured off the supernatant, and stored in the -20 freezer.

2.4 PROTEIN PURIFICATION

The cell pellet was thawed and resuspended in lysis buffer containing 25 mM MES pH 6.5, 50 mM NaCl, 10 mM beta-mercaptoethanol (BME), and 0.5 mM EDTA. Immediately before cell lysis, PMSF (serine protease inhibitor) was added to the lysate to a final concentration of 0.5 mM. The cells were then lysed by running through a microfluidizer. The lysate was then clarified by centrifugation at 12,000 rpm in an SS34 rotor (Sorvall) for 50 minutes. After the spin, the supernatant was added to a chilled glass beaker that was placed on a turn table. While being agitated, 10% streptomycin sulfate solution was slowly added to the lysate so that the DNA bound to protein would precipitate
The streptomycin was added slowly in half mL increments until the solution took on an opaque color. This solution was then centrifuged at 12,000 rpm in an SS34 rotor for 50 minutes. After the spin, the soluble lysate was filtered using a 0.8μm filter.

The lysate was then loaded onto a 15 mL SP sepharose (GE Healthcare) column equilibrated with lysis buffer. The column was washed with 200 mL of lysis buffer. Protein was eluted with a salt gradient of 50-500 mM NaCl in lysis buffer over twenty column volumes. Samples from every fifth fraction were collected and run on a 12.5% SDS-PAGE gel to find which fractions contained protein. Those fractions were pooled and concentrated to a total volume of about 3 mL. The protein solution was filtered with a 0.2μm filter to remove any large particles. The lysate was then loaded onto a Superdex 200 26/60 size exclusion column (GE Healthcare) equilibrated with a column running buffer consisting of 25mM Tris pH 7.4, 50mM NaCl, and 1mM DTT. The corresponding chromatograph indicated which fractions the protein of interest came off in and those were pooled and concentrated in centrifugal concentrators (Millipore) to a final concentration of 15 mg/mL. The protein was flash frozen in liquid nitrogen and stored at -80°C in 25μL aliquots.

### 2.5 Complex Formation and Crystallization

A construct corresponding to residues 39-363 of NF-κB p50 homodimer was expressed in *E. coli* and further steps were taken to purify the protein. The protein was then concentrated to 15 mg/mL. The 16mer overhang: p50 homodimer complex stock was prepared by mixing 25μL of p50 homodimer at15mg/mL with 3μL of 16mer overhang (2.8 mM) DNA. The mixture was then incubated on ice for 10 minutes.

To form the DNA: protein complex, double stranded DNA was added to protein in a 2:1 or 1:1 ratio, respectively. For the crystallization experiments, the hanging drop vapor diffusion method was employed, as well as sitting drop vapor diffusion. In hanging drop vapor diffusion, siliconized glass cover slips cover glass and 24-well VDX plates (Hampton Research). A thin bead of silicone-based vacuum grease was applied to each well. All trays were set up and incubated at room temperature. Typically, 500-1000 μL of crystallization solution was placed into each well. In hanging drop vapor screens, 1 μL of solution was placed onto a siliconized glass coverslip. 1μL of protein complex was then mixed with the 1μL of reagent to form the final 2μL drop. The coverslip was then inverted over the well
containing the reagent and sealed completely. Sitting drop vapor diffusion works in a similar fashion to hanging drop but uses TTP Labtech’s mosquito Crystal. Instead of hanging on the coverslip, the drop sits in the well slightly above the reservoir solution. The mosquito allows you to create several multiple component drops per well in 96-well hanging drop set-ups, allowing you to assess different protein concentrations, ligands or complexes at the same time. Crystal trays were monitored daily by visual inspection through a stereomicroscope for formation of crystals.

The initial screening was performed by random screening of Hampton Research solutions for crystal growth. These consist of a variety of different buffers, precipitants, and salts. This gives information on how different reagents affect crystal growth. Once an interesting hit is obtained from the randomized crystal screens, the researcher can then produce their own reagents to optimize the given conditions of the hit to improve crystal growth and quality.

### 2.6 X-Ray Diffraction Experiments

Once a protein crystal has grown, it can then be harvested for use in x-ray experiments. The crystal of interest is fished out of the drop by Nylon loops (Hampton Research). The crystal is then placed in a cryo protectant reagent or can be flash frozen directly. The nylon loops containing the crystals are then stored in liquid nitrogen. Crystals were tested for diffraction either in the SDSU Macromolecular X-ray Crystallography Facility or the Advanced Photon Source station 24-ID-C at Argonne National Laboratory.

### 2.7 Structural Determination

Molecular replacement was used to solve the phase problem. Molecular replacement uses a previously solved structure to predict what a similar, unknown protein structure will look like. The model used for molecular replacement was the p50 homodimer bound to IL-6 DNA solved by Norman Zhu (PDB ID: 4RTV). Both crystals were from the same construct of p50, 39-363, so Zhu’s model was ideal for Molecular Replacement and no residues had to be removed. p50 has two domains, the C terminus, or dimerization domain, and the N-terminus domain, or DNA binding domain. Since it is a homodimer, the model had a total of
4 domains, as well as the DNA which were all put into Phaser software program in order to obtain one clear solution (Emsley and Cowtan, 2004).

### 2.8 Model Building and Refinement

Model building and refinement were accomplished by using the software Coot and Phoenix (Adams et al., 2010). Molprobity is part of the Phoenix software and was used to improve the geometry of the structure (Chen et al., 2010). Changes to the model that resulted in a lower $R$-free value were kept and those that increased the $R$-free were undone. $R$-free is the measure of error in the model, so the lower the $R$-free, or error, the better the model is (McCoy et al., 2007). Refinement was concluded when the $R$-free value reached 34.6%.
CHAPTER 3

RESULTS

3.1 PROTEIN EXPRESSION AND PURIFICATION

Expression and purification of murine p50 (39-363) homodimer in *Escherichia coli* was performed to achieve large stocks of the protein of interest. This construct contains the entire Rel homology region (RHR), DNA binding, and dimerization segments of the protein. Norman Zhu created a protocol for the purification of the p50 homodimer Rel homology region after performing the expression and purification himself numerous times. It is essential to prepare large protein stocks for crystallization studies because variations between protein preps can affect crystallization.

Two 4-liter cultures were used for every protein preparation. Ion exchange chromatography was used as part of the purification; specifically the SP Sepharose cation exchange column. Sepharose is a cross-linked beaded form of agarose. A cation exchange column catches positive ions and p50 is positive so the column catches it. Increased salt concentration was used to elute p50 and then an SDS-PAGE gel was used to find out where it eluted. p50 is 73 kDa. During ion exchange chromatography, samples were taken of the cell lysate, flow through, wash, and then every 5th fraction during the elution stage. These samples were run on an SDS-PAGE gel. The gel indicates which fractions contain the eluted protein (Figure 3.1A).

The fractions containing protein were subsequently pooled and concentrated with 2 Centri-Prep 30K concentrators. After concentrating the protein down to around 3 mL total volume, p50 was subsequently purified by FPLC, fast protein liquid chromatography, which is a type of size exclusion chromatography (Figure 3.1B). Fractions containing protein were concentrated to 15 mg/mL. Samples were stored in 25μL aliquots, flash frozen, and stored at -80°C. With x-ray crystallography it is essential to obtain a high yield of really pure protein in order to grow the protein crystal of interest.
Figure 3.1. Purification of NF-κB p50. (A) SDS-PAGE gel of ion exchange purified p50 protein. (B) Chromatograph from size exclusion chromatography using Superdex 200 26/60. The blue curve represents protein, which absorbs at 280nm. The x-axis is the elution volume in mL and the y-axis is absorbance at 280 nm in mAU.
3.2 DNA Oligomer Design and Purification

Previous studies determined that p50 can bind palindromic DNA containing three consecutive guanines on each 5' end and thus repress transcription. Dr. Tom Huxford and I designed single strand DNA oligomers and ordered them from IDT. Oligos were purified using a source Q anion exchange column and eluted with Tris and NaCl. Fractions containing pure oligo were concentrated to 2-3 mM (Figure 3.2). Complementary oligos were combined, annealed, and concentrated to 2-3 mM. This protocol is necessary for the successful purification of oligos and to obtain clean oligos for crystallographic studies.

3.3 Crystallization & X-Ray Diffraction

Figure 2.1 illustrates the different DNA oligos screened for crystallization. Hampton Research macromolecular crystallization kits were used in the screening process. PEG ion screen 1 & 2, Index screens, crystal screen 1 & 2, and crystal screen cryo were tested. These are a mixture of different crystallization reagents. Among the conditions screened, there were a variety of results seen. Clear drops and precipitation were most commonly seen. DNA and p50 were combined in 1:1 ratio to form the protein/DNA complex used in the Hampton screening.

Only the 16mer over hang DNA crystallized (Figure 3.3). The p50 homodimer:16mer overhang complex was crystallized by sitting drop vapor diffusion (reservoir solution: 4% v/v Tacsimate at pH 6.0 and 12% w/v Polyethylene Glycol). The drop was set up using 100 nL of p50 homodimer: 16mer overhang mixture with 100 nL of reservoir solution over 59.9 nL of reservoir solution (60 nL-100 nL). It took one week to grow at 20°C. Tacsimate is an organic salt consisting of malonic acid, ammonium citrate tribasic, succinic acid, DL-malic acid, sodium acetate trihydrate, sodium formate, and ammonium tartrate dibasic. The crystal was grown using UCSD’s Mosquito crystal robot to test different reservoir solution conditions.
Figure 3.2. DNA Purification of κB DNA Oligos. The x-axis corresponds to elution volume in mL and the y-axis is in mAU. (A) Chromatograph from size exclusion chromatography of 16mer OHT (B) Chromatograph of 16mer OHA DNA Purification.
TTP Labtech’s Mosquito crystallization robot also allows the user to create several multiple component drops per well in 96-well hanging drop set-ups, allowing you to assess different protein concentrations, ligands or complexes at the same time. The Mosquito uses the sitting drop vapor diffusion method. The crystal was immersed in a cryo protectant consisting of 40% glycerol and reservoir solution, before being flash frozen in liquid nitrogen. The crystal was around 140 microns across the top in length. Because only one crystal grew, it was impossible to see trends or patterns in crystallization growth. The crystal was tested for diffraction at the Advanced Photon Source station 24-ID-C at Argonne National Laboratory. The crystal diffracted to 3 Å. Figure 3.4 shows the crystal data diffraction pattern with XDS indexed intensities used to determine the structure factors. Data collection and diffraction resolution is a critical step in x-ray crystallography and if the resolution is not high enough, it is impossible to solve the structure correctly.
Figure 3.4. Diffraction Image. The crystal was tested for diffraction at the Advanced Photon Source station 24-ID beam C at Argonne National Laboratory. Shown here is the crystal data diffraction pattern with XDS indexed intensities used to determine the structure factors.

X-ray diffraction data can be assessed both in terms of the quantity and quality of the measured intensities. Some quantifiable measures of x-ray diffraction data include the data resolution, the completeness of collected x-ray diffraction intensities, the signal-to-noise ratio, and the residual (R) factor (Otwinowski and Minor, 1997). Resolution is a measure of the distance between parallel planes that slice through the unit cell and, consequently, the level of detail with which electron density maps can be calculated. Smaller values correlate with smaller spacings, which are typically measured in Å and that correlate with higher sampling of the unit cell contents and thus higher resolution. Higher resolution data (lower values of resolution in units of Å) generally permits better visualization of atoms and bonds in electron density maps. Hydrogen bond distances and well defined water molecules can be viewed at resolutions of better than 2.7 Å. Carbon-carbon covalent bonds can usually be seen at resolutions better than 1.5 Å. The p50:idealized kB DNA complex structure has a resolution of 3 Å which is on the lower side of medium in terms of resolution.

Data completeness refers to the total number of reflection intensities measured as a fraction of the theoretical limit at any particular resolution. Obviously, the higher the completeness the better. Although electron density maps are often generated through Fourier synthesis of intensities from data sets that are not complete, the systematic absence of
segments of data can generate problems in electron densities maps that make model building and refinement a challenge.

The signal-to-noise ratio compares the average intensity (I) for each of the measured reflections against the background intensity (σ) of that position in the diffraction pattern when there is no reflection to measure. As with any analytical technique, the higher the value of I/σ the more confidence one has that the actual value has been accurately measured. Reflection intensity generally decreases as a function of resolution so that the highest resolution reflections always exhibit significantly lower values of I/σ than do lower resolution reflections. As a rule of thumb, collection of complete data with an I/σ greater than 2.0 in the highest resolution data can be confidently used in the calculation of electron density maps and modeling of x-ray crystal structures. This rule, like all other statistical values, should always be checked against the other parameters rather than being blindly followed. For example, in completing data collection one might collect a highly redundant data set in which the same measurements were made over and over again. In such cases, it might make sense to include data that were identified with average intensities that are only slightly higher than the background.

The final measure of x-ray diffraction data quality is the residual or R-factor. This compares the reproducibility of measurements of the same reflection intensity. In this case, the "same" reflection could mean the actual reflection is measured multiple times or another reflection that is identical due to symmetry in the crystal. In the process of x-ray diffraction data collection, each reflection intensity is typically measured numerous times allowing for a comparison of how well multiple measures of intensity agree with one another. The calculation is the sum of all the differences between intensities measured and the average value for that reflection as a percentage of the average value. This makes it so that perfect measurement of the exact same value of intensity would give a value of 0% whereas, purely random numbers would generate value somewhere around 60%. Lower values for the overall residual of all data, which is referred to as Rsymm, correlates with data that are of high quality while higher values of Rsymm can mean that a crystal was improperly handled or slipped during data collection or was destroyed by x-radiation. Rsymm values should be evaluated in consideration with the redundancy of data and overall symmetry of the crystal as high redundancy data often exhibits higher values of Rsymm but can represent better average
values of intensity due to the greater number of measurements made. So, as this discussion
illustrates, no one single parameter is sufficient to assess x-ray reflection data quality.
Rather, through an analysis of these various values one can gain a quantitative and qualitative
assessment of x-ray diffraction data quality.

3.4 Ideal P50 Homodimer Structure

The x-ray crystal structure of NF-κB p50 (39-363) homodimer bound to ideal 16 base
pair overhang κB DNA was solved by molecular replacement and refined against diffraction
data that was complete to a resolution of 3.0 Å. All the refinement statistics associated with
the structure are listed in Table 3.1.

The overall structure is analogous to the p50 homodimer structure published at
Harvard in 1995 (Müller and Harrison, 1995). The p50 monomer is separated into two
domains, the C terminus, or dimerization domain, and the N terminus, or DNA binding
domain. The C terminal domain contains a core beta sandwich structure. The N-terminal
Domain has a variable length alpha helical segment that forms strong contacts with the minor
groove near the center of the κB element. The construct of p50 is 39-363, and the
dimerization domain consists of residues 245-363, and the N-terminus domain contains
residues 39-245. The RHR found at the amino terminus of these proteins, is responsible for
protein dimerization, DNA binding, and nuclear localization (Müller and Harrison, 1995).

Since the structure is a homodimer, it takes on a butterfly appearance with the DNA
double helix in the middle (Figure 3.5). There is a flexible linker between the N and C
terminus domains. The linker is what allows the p50 homodimer to dock onto the DNA and
vary the spacing between the two N terminus domains so that there can be different half site
spacing. This is what allowed p50 to adopt an 11 base pair spacing preference (Figure 3.6).
### Table 3.1. Data Collection and Refinement Statistics

<table>
<thead>
<tr>
<th></th>
<th>p50:Ideal κB DNA complex</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data collection</strong></td>
<td></td>
</tr>
<tr>
<td>X-ray source</td>
<td>APS 24-ID-C</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.97920</td>
</tr>
<tr>
<td>Space group</td>
<td>P2₁</td>
</tr>
<tr>
<td>Unit Cell (Å)</td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>47.15</td>
</tr>
<tr>
<td>b</td>
<td>146.45</td>
</tr>
<tr>
<td>c</td>
<td>67.69</td>
</tr>
<tr>
<td>beta (°)</td>
<td>98.62</td>
</tr>
<tr>
<td>Complexes/asymm. unit</td>
<td>1</td>
</tr>
<tr>
<td>Resolution range (Å)¹</td>
<td>146.45-3.00 (3.16-3.00)</td>
</tr>
<tr>
<td>R&lt;sub&gt;sym&lt;/sub&gt; (%)</td>
<td>9.4 (63.1)</td>
</tr>
<tr>
<td>Observations</td>
<td>61,285</td>
</tr>
<tr>
<td>Unique Reflections</td>
<td>17,962</td>
</tr>
<tr>
<td>Redundancy</td>
<td>3.4 (3.5)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>98.6 (99.1)</td>
</tr>
<tr>
<td>&lt;I/σ&gt;</td>
<td>10.1 (2.0)</td>
</tr>
<tr>
<td><strong>Structure solution by MR</strong></td>
<td></td>
</tr>
<tr>
<td>Probe</td>
<td>4RTV</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>4-10</td>
</tr>
<tr>
<td>Z-score (from Phaser)</td>
<td>20.7</td>
</tr>
<tr>
<td>LLG (from Phaser)</td>
<td>1828</td>
</tr>
<tr>
<td><strong>Refinement</strong></td>
<td></td>
</tr>
<tr>
<td>Resolution Range (Å)</td>
<td>73.25-3.00(3.16 -3.00)</td>
</tr>
<tr>
<td>Unique Reflections</td>
<td>17933</td>
</tr>
<tr>
<td># of Atoms</td>
<td>5597</td>
</tr>
<tr>
<td>R-work</td>
<td>27.7(36.7)</td>
</tr>
<tr>
<td>R-free</td>
<td>34.6(43.5)</td>
</tr>
<tr>
<td>Geometry</td>
<td></td>
</tr>
<tr>
<td>R.m.s. bond lengths (Å)</td>
<td>0.004</td>
</tr>
<tr>
<td>R.m.s. bond angles (°)</td>
<td>0.911</td>
</tr>
<tr>
<td>Mean B (Å²)</td>
<td>71.57</td>
</tr>
<tr>
<td>Clashscore</td>
<td>15.53</td>
</tr>
<tr>
<td>Ramachandran plot</td>
<td></td>
</tr>
<tr>
<td>Favored</td>
<td>85.81</td>
</tr>
<tr>
<td>Disallowed</td>
<td>2.90</td>
</tr>
<tr>
<td>MolProbity Score</td>
<td>2.47</td>
</tr>
</tbody>
</table>
Figure 3.5. Ribbon diagram of p50 (39-363) Rel Homology Region. (A) p50 subunits are represented in green. The multi-colored barrel in the middle represents the DNA. The C represents the dimerization domain and the N represents the DNA binding domain. (B) The view rotated 90°.
Figure 3.6. Protein-DNA interactions of the p50 homodimer bound to various promoters. From left to right, the Yale structure of p50 homodimer bound to a symmetrized κB site and 12 base pair spacing between the p50 subunits, Harvard structure bound to MHC class I enhancer site and 11 base pair spacing, p50 bound to IL-6 promoter and 11 base pair spacing, p50 bound to the NGAL promoter and 12 base pair spacing, and the structure I solved with the p50 homodimer bound to “idealized” κB DNA and 11 base pair spacing. The amino acids histidine, and two arginine’s are labeled on where they contact the DNA promoters.
CHAPTER 4

DISCUSSION

The objective of this project was to obtain a crystal structure of p50 bound to a theoretical palindromic κB DNA site in order to determine preferred binding mode of p50. Then I wanted to determine how this correlates with what is seen in other NF-κB binding sites. There have been numerous p50 homodimer structures published to date. The first p50 homodimer structure was solved in 1995 by researchers at Yale. The dimer was bound to a symmetrical κB site and had 12 base pair spacing between the p50 subunits (Ghosh et al., 1995). Around the same time, researchers at Harvard published a model of p50 bound to an MHC class I enhancer site but this model had an 11 base pair spacing between the two p50 subunits. In the Huxford lab, Norman Zhu solved two p50 structures. p50 bound to the IL-6 promoter had 11 base pair spacing and p50 bound to the NGAL promoter had 12 base pair spacing. So with two structures showing an 11 base pair preference and two structures showing a 12 base pair preference, there needed to be a tie breaker to determine the preferred binding mode of p50. Thus, DNA was designed with 4 adjacent Guanines on the 5' ends of each strand. This would allow p50 to adopt either the 11 or 12 base pair spacing between the two N terminus domains.

p50 bound to this “idealized” κB DNA was crystallized (Figure 3.3) and the structure was solved using molecular replacement. Molecular replacement was used only for the p50 homodimer structure but not for the DNA. As you can see in Figure 4.1, there is electron density for DNA and it looks like a double helix. This is important because the data shows there is DNA in the model, as opposed to us manually inputting the DNA. If you zoom in on the electron density map, you can see the two arginine residues and one histidine making contact with the three adjacent guanines (Figure 4.2). Now it can be confidently stated that we have crystallized p50 homodimer bound to our ideal κB DNA.
After some refinement, a model was built (Figure 3.5 and 4.3) and p50 adopted an 11 base pair spacing. The amino acids histidine, and two arginines are labeled on where they contact the DNA promoters (Figure 3.6). We now have a better understanding of the p50 homodimer preferred binding conformation. If given a choice, it prefers to bind with 11 base pair spacing. This correlates with what was shown in the Harvard structure solved in 1995.
In order to learn more about the site where p50 is binding on DNA, affinity binding studies are required. Microscale thermophoresis allows the user to determine protein: protein or protein: DNA dissociation constants. This measurement of equilibrium binding affinity is accomplished by using small volumes of protein:protein or protein:DNA complexes drawn up into capillaries that are heated by an infrared laser. The instrument then measures the movement of fluorescently tagged protein or DNA and calculates the dissociation constant. I would like to determine how tightly p50 is bound to κB DNA and how this correlates with κB DNA sites that IκBζ influences by using microscale thermophoresis to demonstrate if the addition of IκBζ acts like unlabeled DNA to pull p50 off the labeled probe or if it generates a ternary complex.
Previous studies conducted by Norman Zhu using electrophoretic mobility shift assays, showed that IκBζ was bound to the C-terminus of the p50 homodimer while the N terminus of p50 was still bound to IL-6 DNA and a new ternary complex was formed and the gene of interest was transcribed (Trinh et al., 2008). Based on these results, I hypothesize that this ternary complex will be confirmed using microscale thermophoresis. The larger complex will show decreased “thermophoresis” compared to p50 homodimer bound to DNA alone because bigger substrates take longer to move, as well as an increase in binding as the concentration of IκBζ is increased.
ACKNOWLEDGEMENTS

I would like to thank Dr. Tom Huxford for guiding and supporting me throughout this whole process.

I would also like to thank Dr. Harris and Dr. Cole for taking the time to serve on my thesis committee.

Thank you to everyone in the Huxford lab, especially Lisa Acuna, Eric Rodgers, Ashlee King, Norman Zhu, and James Caldwell.

This work is based upon research conducted at the Northeastern Collaborative Access Team beamlines, which are funded by the National Institute of General Medical Sciences from the National Institutes of Health (P41 GM103403). The Pilatus 6M detector on 24-ID-C beam line is funded by a NIH-ORIP HEI grant (S10 RR029205). This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357.
REFERENCES


